

# nature

**STEM CELL THERAPY**

Ask the patient

**HENRIKSS**

The sword for  
a stable nation

**DIAMOND EXPECTS**

Marine life  
NATC imaging



## SPECIATION IN COLOUR

A textbook example of  
evolution in action

**HAPLODIPLOIDY**  
Graham & Pavesi

# A question of balance

The turmoil in the financial markets could lead to severe cost-cutting by governments, but US politicians would do well to note the benefits of continued support for clean energy and climate policies.

As *Nature* went to press, a US\$700-billion rescue plan for the collapsing financial sector remained in limbo in the US Congress. Some such deal may yet be passed. If it is, it still may not be enough to stop the global economy from sliding into recession. Faced with the prospect of fresh outlays and declining revenue, the US government may soon be looking for ways to tighten its belt, a situation likely to be echoed in Europe and beyond. For science and technology, this could mean less money for basic research, education and clean energy, and could pose fresh threats to the long-promised climate legislation in the United States.

Those who favour cutbacks — or, in the case of climate regulations, not moving forward — will say such activities cannot be afforded. In some cases, they may well be right; scientists may have to gird themselves for flat budgets into the foreseeable future, and set their priorities accordingly. But there is a danger that the debate will be framed entirely in terms of costs, with no consideration of the benefits. Investment in areas such as research, education and clean energy are part of the foundations for long-term prosperity.

The good news in the United States is that both leading presidential candidates have made this connection, especially in the energy and climate arena. Neither the Republican candidate John McCain nor the Democratic candidate Barack Obama is selling climate regulation as an expensive moral obligation to the environment. Instead, both speak of the benefits of 'green-collar' jobs and energy security.

There are concerns that McCain might succumb to pressure from the far-right of his party and back away from his pledge to curb greenhouse-gas emissions through a cap-and-trade programme. But he has yet to do so. His advisers continue to advocate hybrid vehicles, for instance, as both a cheaper and cleaner alternative to the internal combustion engine and a way to make the United States less dependent on increasingly expensive foreign oil.

Obama has gone further, integrating energy and climate policy with his plan for revitalizing the US economy. He argues that green jobs tend to be domestic jobs, which means energy security goes hand-in-hand with economic development. This might be

dismissed as overly optimistic, given that the transition to clean energy won't be cheap, but there is little doubt that new industries will eventually rise in place of the old ones. In last week's first presidential debate, Obama also endorsed solid investments in science and technology generally.

It's refreshing to see that this political realignment has also taken hold in Congress. Efficiency and conservation, frequently played down as feel-good measures in the past, are now seen as critical components of the energy equation.

The notion that the government can use its purchasing power to advance the development of clean vehicles has been heralded as a way of increasing energy security while addressing the long-term threat of climate change.

Such ideas, if implemented, will drive new investment in the years to come. It will take time for the world's financial institutions to rebuild

themselves following the implosion on Wall Street. But the fundamental need to create a more sustainable energy infrastructure to power the globe will be as strong as ever. Eventually the market will respond to that opportunity.

It would be naive to assume that progress on these issues will be easy, even with vigorous leadership. Private investment in new technology has increased significantly in recent years — but so have global greenhouse-gas emissions. This makes it all the more important that Congress and the international community move quickly to establish a solid and predictable climate regulatory framework to carry the world beyond the Kyoto Protocol. The current market turmoil is due in part to a continuing crisis of confidence, so a little regulatory certainty on greenhouse gases might be welcome. It won't solve the financial crisis by itself, but it would help businesses, financial institutions and funding agencies place their bets on the future. ■

**"The fundamental need to create a more sustainable energy infrastructure to power the globe will be as strong as ever, and the market will respond to that opportunity."**

## Life after Zerhouni

The next NIH director must juggle stagnant budgets, unhappy grantees and investigative lawmakers.

The imminent departure of Elias Zerhouni as director of the US National Institutes of Health (NIH) in Bethesda, Maryland, leaves large shoes to fill. Zerhouni, who announced last week he will quit his post by the end of October (see page 570), managed the agency with a blend of vision, toughness and dedication even

as it faced stagnating funding, ethical uproars and an explosion of knowledge in biomedical research. It is to his credit that he leaves the agency with far more friends than enemies, and with a well-earned reputation as a public servant who tirelessly maintained his integrity during the administration of President George W. Bush.

Whoever follows Zerhouni faces unenviable challenges. Stem-cell research remains stymied by an outdated presidential policy. The conflict-of-interest scandal continues as Senator Charles Grassley (Republican, Iowa) has reported troubling instances of extramural NIH researchers failing to report five- and six-figure payments from drug companies that could benefit from their research. And with

# A question of balance

The turmoil in the financial markets could lead to severe cost-cutting by governments, but US politicians would do well to note the benefits of continued support for clean energy and climate policies.

As *Nature* went to press, a US\$700-billion rescue plan for the collapsing financial sector remained in limbo in the US Congress. Some such deal may yet be passed. If it is, it still may not be enough to stop the global economy from sliding into recession. Faced with the prospect of fresh outlays and declining revenue, the US government may soon be looking for ways to tighten its belt, a situation likely to be echoed in Europe and beyond. For science and technology, this could mean less money for basic research, education and clean energy, and could pose fresh threats to the long-promised climate legislation in the United States.

Those who favour cutbacks — or, in the case of climate regulations, not moving forward — will say such activities cannot be afforded. In some cases, they may well be right; scientists may have to gird themselves for flat budgets into the foreseeable future, and set their priorities accordingly. But there is a danger that the debate will be framed entirely in terms of costs, with no consideration of the benefits. Investment in areas such as research, education and clean energy are part of the foundations for long-term prosperity.

The good news in the United States is that both leading presidential candidates have made this connection, especially in the energy and climate arena. Neither the Republican candidate John McCain nor the Democratic candidate Barack Obama is selling climate regulation as an expensive moral obligation to the environment. Instead, both speak of the benefits of 'green-collar' jobs and energy security.

There are concerns that McCain might succumb to pressure from the far-right of his party and back away from his pledge to curb greenhouse-gas emissions through a cap-and-trade programme. But he has yet to do so. His advisers continue to advocate hybrid vehicles, for instance, as both a cheaper and cleaner alternative to the internal combustion engine and a way to make the United States less dependent on increasingly expensive foreign oil.

Obama has gone further, integrating energy and climate policy with his plan for revitalizing the US economy. He argues that green jobs tend to be domestic jobs, which means energy security goes hand-in-hand with economic development. This might be

dismissed as overly optimistic, given that the transition to clean energy won't be cheap, but there is little doubt that new industries will eventually rise in place of the old ones. In last week's first presidential debate, Obama also endorsed solid investments in science and technology generally.

It's refreshing to see that this political realignment has also taken hold in Congress. Efficiency and conservation, frequently played down as feel-good measures in the past, are now seen as critical components of the energy equation.

The notion that the government can use its purchasing power to advance the development of clean vehicles has been heralded as a way of increasing energy security while addressing the long-term threat of climate change.

Such ideas, if implemented, will drive new investment in the years to come. It will take time for the world's financial institutions to rebuild

themselves following the implosion on Wall Street. But the fundamental need to create a more sustainable energy infrastructure to power the globe will be as strong as ever. Eventually the market will respond to that opportunity.

It would be naive to assume that progress on these issues will be easy, even with vigorous leadership. Private investment in new technology has increased significantly in recent years — but so have global greenhouse-gas emissions. This makes it all the more important that Congress and the international community move quickly to establish a solid and predictable climate regulatory framework to carry the world beyond the Kyoto Protocol. The current market turmoil is due in part to a continuing crisis of confidence, so a little regulatory certainty on greenhouse gases might be welcome. It won't solve the financial crisis by itself, but it would help businesses, financial institutions and funding agencies place their bets on the future. ■

**"The fundamental need to create a more sustainable energy infrastructure to power the globe will be as strong as ever, and the market will respond to that opportunity."**

## Life after Zerhouni

The next NIH director must juggle stagnant budgets, unhappy grantees and investigative lawmakers.

The imminent departure of Elias Zerhouni as director of the US National Institutes of Health (NIH) in Bethesda, Maryland, leaves large shoes to fill. Zerhouni, who announced last week he will quit his post by the end of October (see page 570), managed the agency with a blend of vision, toughness and dedication even

as it faced stagnating funding, ethical uproars and an explosion of knowledge in biomedical research. It is to his credit that he leaves the agency with far more friends than enemies, and with a well-earned reputation as a public servant who tirelessly maintained his integrity during the administration of President George W. Bush.

Whoever follows Zerhouni faces unenviable challenges. Stem-cell research remains stymied by an outdated presidential policy. The conflict-of-interest scandal continues as Senator Charles Grassley (Republican, Iowa) has reported troubling instances of extramural NIH researchers failing to report five- and six-figure payments from drug companies that could benefit from their research. And with

nothing but flat funding for the foreseeable future, NIH-supported labs are being squeezed nationwide. A new generation of academic scientists is being imperilled as many head for jobs in industry or elsewhere rather than face the daunting odds of ever landing an NIH grant.

All of this is occurring in the post-human-genome era with a knowledge base expanding at warp speed. Improved understanding and treatment of diseases have never been so tantalizingly close. So what qualities should the next president seek in a new NIH director? Three are key.

First, despite the fact that some two-thirds of the agency's budget is spent on basic research, the next director should be someone who understands, and is committed to, translating discoveries to the bedside. Zerhouni, a radiologist, did much to advance this agenda, although it was not his idea; the agency's mission statement makes it clear that the NIH is devoted to "science in pursuit of fundamental knowledge... and the application of that knowledge to extend healthy life and reduce the burdens of illness and disability". The agency's next director should not throw money willy-nilly at translational research; accountability is vital as such work goes forward. But taxpayers who

invest US\$29 billion annually in the NIH deserve to see their lives and health improved because of it.

Second, the next director should be a gifted communicator who can speak with ease to the NIH's scientific constituency, to Congress and to the public. Translating complex research into terms meaningful to the public and to lawmakers is a crucial skill, especially as the NIH seeks its share of an ever-more-constrained federal budget.

Third, the next director should be an able manager willing to make and stick to tough decisions in times of ethical and financial stress. Although the ranks of current and former directors of the agency's 27 component institutes contain many amply qualified candidates for the top job, it may be worth reaching outside NIH circles for a candidate not beholden to long-time peers in Bethesda. Zerhouni, who came from the Johns Hopkins School of Medicine in Baltimore, Maryland, showed that this strategy can work well.

Overall, a director should be chosen with appropriate speed. Allowing the NIH's top post to sit vacant for months or years — as Bush did when he took more than two years to nominate Zerhouni — could do serious damage to the agency at a time when bold leadership is vital. ■

## An end to secrecy

China's continuing openness on HIV is a welcome development and a model for other nations.

**A**s part of a special collection of articles on HIV, this week's issue contains a Feature by Linqi Zhang of Tsinghua University in Beijing and his colleagues on the status of HIV in southern China (see page 609). Their conclusions are alarming: HIV prevalence is no longer confined to high-risk groups such as those who inject themselves with drugs, but is now seeping into the general population. Some of the most rapid increases are among men in same-sex relationships. Moreover, the findings confirm what veteran outside-observers of China and those concerned with HIV globally have long suspected: patterns of infection in southern China are similar to those in other developing countries — especially those experiencing large-scale migration from rural areas to cities, which provides men and women with more opportunities for sex.

The good news, however, is that China is doing more to make its AIDS statistics available. Traditionally, China has controlled access to such information very tightly. After the first AIDS cases were reported in the 1980s, for example, it took the Chinese government more than a decade to acknowledge publicly that the epidemic even existed. But during the SARS epidemic of 2002–03, the government's secrecy drew the outrage of Chinese journalists and non-governmental organizations alike; the resulting outcry led to a change in official attitudes.

The work of Zhang and his colleagues illustrates just how radical this change has been. Although the study was led by scientists inside China, the group included a leading US-based researcher, David Ho of the Rockefeller University in New York. The international team had

full access to data supplied by government authorities — the results of tests from 3.2 million blood samples. And the authorities apparently made no attempt to control or influence the authors' opinions.

Giving outsiders access to sensitive public health information would have been unthinkable in China even a few years ago — just as it is in many Western countries even now. But then, China is slowly becoming more comfortable with the idea that all of society will benefit by sharing data and knowledge with others. Some of this transparency can be traced back to 1972 and the landmark meeting between US President Richard Nixon and China's Chairman Mao Zedong. As noted by the historian Margaret MacMillan, author of the 2007 book *Nixon and Mao: The Week that Changed the World*, China had a very pragmatic reason for the rapprochement: it needed access to US technology. That opening was greatly expanded by Mao's successor, Deng Xiaoping. Deng accelerated scientific contacts with the rest of the world, sent hundreds of thousands of Chinese students to study in Western universities, and in 1987 hosted a landmark scientific conference in Beijing between China and the international community (see page 598).

Of course, opening up on information is not the same as successfully controlling the spread of infection. Much more needs to be done if the government is to meet its self-imposed target of limiting the total number of cases of HIV infection to 1.5 million by 2010. Nonetheless, transparency is an essential first step. There are the many nations — in North Africa and the Middle East, for example — where public discussion of HIV and its causes is still not as open as it could be.

China was once in a similar position — but it changed. There are many good reasons why others should follow suit. ■

**"Giving outsiders access to sensitive public health information would have been unthinkable in China even a few years ago."**



nothing but flat funding for the foreseeable future, NIH-supported labs are being squeezed nationwide. A new generation of academic scientists is being imperilled as many head for jobs in industry or elsewhere rather than face the daunting odds of ever landing an NIH grant.

All of this is occurring in the post-human-genome era with a knowledge base expanding at warp speed. Improved understanding and treatment of diseases have never been so tantalizingly close. So what qualities should the next president seek in a new NIH director? Three are key.

First, despite the fact that some two-thirds of the agency's budget is spent on basic research, the next director should be someone who understands, and is committed to, translating discoveries to the bedside. Zerhouni, a radiologist, did much to advance this agenda, although it was not his idea; the agency's mission statement makes it clear that the NIH is devoted to "science in pursuit of fundamental knowledge... and the application of that knowledge to extend healthy life and reduce the burdens of illness and disability". The agency's next director should not throw money willy-nilly at translational research; accountability is vital as such work goes forward. But taxpayers who

invest US\$29 billion annually in the NIH deserve to see their lives and health improved because of it.

Second, the next director should be a gifted communicator who can speak with ease to the NIH's scientific constituency, to Congress and to the public. Translating complex research into terms meaningful to the public and to lawmakers is a crucial skill, especially as the NIH seeks its share of an ever-more-constrained federal budget.

Third, the next director should be an able manager willing to make and stick to tough decisions in times of ethical and financial stress. Although the ranks of current and former directors of the agency's 27 component institutes contain many amply qualified candidates for the top job, it may be worth reaching outside NIH circles for a candidate not beholden to long-time peers in Bethesda. Zerhouni, who came from the Johns Hopkins School of Medicine in Baltimore, Maryland, showed that this strategy can work well.

Overall, a director should be chosen with appropriate speed. Allowing the NIH's top post to sit vacant for months or years — as Bush did when he took more than two years to nominate Zerhouni — could do serious damage to the agency at a time when bold leadership is vital. ■

## An end to secrecy

China's continuing openness on HIV is a welcome development and a model for other nations.

**A**s part of a special collection of articles on HIV, this week's issue contains a Feature by Linqi Zhang of Tsinghua University in Beijing and his colleagues on the status of HIV in southern China (see page 609). Their conclusions are alarming: HIV prevalence is no longer confined to high-risk groups such as those who inject themselves with drugs, but is now seeping into the general population. Some of the most rapid increases are among men in same-sex relationships. Moreover, the findings confirm what veteran outside-observers of China and those concerned with HIV globally have long suspected: patterns of infection in southern China are similar to those in other developing countries — especially those experiencing large-scale migration from rural areas to cities, which provides men and women with more opportunities for sex.

The good news, however, is that China is doing more to make its AIDS statistics available. Traditionally, China has controlled access to such information very tightly. After the first AIDS cases were reported in the 1980s, for example, it took the Chinese government more than a decade to acknowledge publicly that the epidemic even existed. But during the SARS epidemic of 2002–03, the government's secrecy drew the outrage of Chinese journalists and non-governmental organizations alike; the resulting outcry led to a change in official attitudes.

The work of Zhang and his colleagues illustrates just how radical this change has been. Although the study was led by scientists inside China, the group included a leading US-based researcher, David Ho of the Rockefeller University in New York. The international team had

full access to data supplied by government authorities — the results of tests from 3.2 million blood samples. And the authorities apparently made no attempt to control or influence the authors' opinions.

Giving outsiders access to sensitive public health information would have been unthinkable in China even a few years ago — just as it is in many Western countries even now. But then, China is slowly becoming more comfortable with the idea that all of society will benefit by sharing data and knowledge with others. Some of this transparency can be traced back to 1972 and the landmark meeting between US President Richard Nixon and China's Chairman Mao Zedong. As noted by the historian Margaret MacMillan, author of the 2007 book *Nixon and Mao: The Week that Changed the World*, China had a very pragmatic reason for the rapprochement: it needed access to US technology. That opening was greatly expanded by Mao's successor, Deng Xiaoping. Deng accelerated scientific contacts with the rest of the world, sent hundreds of thousands of Chinese students to study in Western universities, and in 1987 hosted a landmark scientific conference in Beijing between China and the international community (see page 598).

Of course, opening up on information is not the same as successfully controlling the spread of infection. Much more needs to be done if the government is to meet its self-imposed target of limiting the total number of cases of HIV infection to 1.5 million by 2010. Nonetheless, transparency is an essential first step. There are the many nations — in North Africa and the Middle East, for example — where public discussion of HIV and its causes is still not as open as it could be.

China was once in a similar position — but it changed. There are many good reasons why others should follow suit. ■

**"Giving outsiders access to sensitive public health information would have been unthinkable in China even a few years ago."**

# RESEARCH HIGHLIGHTS

## Vampire genes

*Naturwissenschaften* doi:10.1007/s00114-008-0446-0 (2008)

The evolution of the common vampire bat, *Desmodus rotundus*, included three rounds of duplication of a gene that encodes a salivary enzyme involved in breaking down blood clots. *Desmodus* laps the blood of mammals. The other vampire bats — *Diaemus youngi*, which also feeds from mammals but prefers bird blood, and *Diphylla ecaudata* (pictured), which sticks to birds — have only one copy of a plasminogen activator gene, find David Liberles of the University of Wyoming in Laramie and his colleagues.

Their genetic analysis corroborates established species relationships. DNA sequencing revealed three alternative versions of the gene in *Diaemus* and four in *Diphylla*. The four gene copies that *Desmodus* expresses lack a section called Kringle 2. Its deletion may have aided a dietary switch to mammalian blood.



N. GORDON/OSF/PHOTOLIBRARY

## GEOSCIENCES

### Carbon crunch

*Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0805382105 (2008)

India's smashing into Asia around 50 million years ago brought changes far beyond the creation of the world's highest mountain range: the continental collision is widely thought to have altered global climate.

Dennis Kent of Rutgers University in Piscataway, New Jersey, and Giovanni Muttoni at the University of Milan in Italy offer particular mechanisms for this. The researchers' model predicts that the carbon-rich sediments on the former ocean floor stopped being subducted and producing carbon dioxide when the landmasses touched.

Meanwhile, India's drift into more humid equatorial climes increased the uptake of the greenhouse gas through greater weathering of silicates in the Deccan traps (pictured below). This could have lowered atmospheric

carbon dioxide enough to prompt the cooling trend in the Middle to Late Eocene.

## ECOLOGY

### Diatoms downsize

*Proc. R. Soc. B* doi:10.1098/rspb.2008.1200 (2008)

Global warming is predicted to be bad for diatoms. Hungry and heavy as plankton go, they are expected to find themselves with fewer nutrients and sink more quickly as temperature gradients, and thus density gradients, grow, increasing the energy needed for mixing.

However, the total volume of diatoms in Lake Tahoe, on the California–Nevada border, did not change between 1982 and 2006, despite a warming in average air temperatures in the Tahoe Basin, report Monika Winder and her co-workers at the University of California, Davis. Instead, average diatom sizes fell from 67 micrometres to 35 micrometres, stemming the mean sinking speed and altering energy transfer through the food web.

## CANCER BIOLOGY

### Ensuring a welcome

*Nature Cell Biol.* doi:10.1038/ncb1794 (2008)

Before travelling to new organs — or metastasizing — some cancers send chemical signals to prepare the target organ for their arrival.

Yoshiro Maru and his colleagues at the Tokyo Women's Medical University in Japan had previously found that primary tumours in mice secrete growth factors that stimulate lung cells to produce chemoattractant proteins. These recruit white blood cells into the lungs, and the resulting inflammation recruits cancer cells to the site.

The team now reports that the chemoattractants involved induce the synthesis of serum amyloid A3 in lung cells. This protein attracts and activates white blood cells, setting up a state of chronic inflammation that facilitates tumour cell invasion. Antibodies against serum amyloid A3 blocked metastasis.

## GEOLOGY

### Primitive petrous

*Science* **321**, 1828–1831 (2008)

A beige outcrop in northern Quebec may be Earth's oldest known crustal rock. Jonathan O'Neil of McGill University in Montreal, Canada, and his colleagues have dated parts of the stone using ratios of neodymium and samarium isotopes, and calculated the oldest section to be 4.28 billion years old. This is 250 million years older than the previous record-holder.

The rocks in question are from the Nuvvuagittuq greenstone belt. This belt had been estimated to be 3.8 billion years old, based on an analysis of zircon crystals. But the stone that O'Neil and his team probed contained no zircons, forcing them to use an alternative method. The outcrop's low levels of neodymium suggest that it formed before Earth's neodymium levels became fixed 4.1 billion years ago.

## MECHANICS

### Slippery when clean

*Phys. Rev. Lett.* **101**, 125505 (2008)

Friction is a familiar force in everyday life, but its nanometre-scale details are obscure. This is because the fundamental mechanisms are subtle and sensitive to contamination,



DINODIA IMAGES/ALAMY



# RESEARCH HIGHLIGHTS

## Vampire genes

*Naturwissenschaften* doi:10.1007/s00114-008-0446-0 (2008)

The evolution of the common vampire bat, *Desmodus rotundus*, included three rounds of duplication of a gene that encodes a salivary enzyme involved in breaking down blood clots. *Desmodus* laps the blood of mammals. The other vampire bats — *Diaemus youngi*, which also feeds from mammals but prefers bird blood, and *Diphylla ecaudata* (pictured), which sticks to birds — have only one copy of a plasminogen activator gene, find David Liberles of the University of Wyoming in Laramie and his colleagues.

Their genetic analysis corroborates established species relationships. DNA sequencing revealed three alternative versions of the gene in *Diaemus* and four in *Diphylla*. The four gene copies that *Desmodus* expresses lack a section called Kringle 2. Its deletion may have aided a dietary switch to mammalian blood.



N. GORDON/OSF/PHOTOLIBRARY

## GEOSCIENCES

### Carbon crunch

*Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0805382105 (2008)

India's smashing into Asia around 50 million years ago brought changes far beyond the creation of the world's highest mountain range: the continental collision is widely thought to have altered global climate.

Dennis Kent of Rutgers University in Piscataway, New Jersey, and Giovanni Muttoni at the University of Milan in Italy offer particular mechanisms for this. The researchers' model predicts that the carbon-rich sediments on the former ocean floor stopped being subducted and producing carbon dioxide when the landmasses touched.

Meanwhile, India's drift into more humid equatorial climes increased the uptake of the greenhouse gas through greater weathering of silicates in the Deccan traps (pictured below). This could have lowered atmospheric

carbon dioxide enough to prompt the cooling trend in the Middle to Late Eocene.

## ECOLOGY

### Diatoms downsize

*Proc. R. Soc. B* doi:10.1098/rspb.2008.1200 (2008)

Global warming is predicted to be bad for diatoms. Hungry and heavy as plankton go, they are expected to find themselves with fewer nutrients and sink more quickly as temperature gradients, and thus density gradients, grow, increasing the energy needed for mixing.

However, the total volume of diatoms in Lake Tahoe, on the California–Nevada border, did not change between 1982 and 2006, despite a warming in average air temperatures in the Tahoe Basin, report Monika Winder and her co-workers at the University of California, Davis. Instead, average diatom sizes fell from 67 micrometres to 35 micrometres, stemming the mean sinking speed and altering energy transfer through the food web.

## CANCER BIOLOGY

### Ensuring a welcome

*Nature Cell Biol.* doi:10.1038/ncb1794 (2008)

Before travelling to new organs — or metastasizing — some cancers send chemical signals to prepare the target organ for their arrival.

Yoshiro Maru and his colleagues at the Tokyo Women's Medical University in Japan had previously found that primary tumours in mice secrete growth factors that stimulate lung cells to produce chemoattractant proteins. These recruit white blood cells into the lungs, and the resulting inflammation recruits cancer cells to the site.

The team now reports that the chemoattractants involved induce the synthesis of serum amyloid A3 in lung cells. This protein attracts and activates white blood cells, setting up a state of chronic inflammation that facilitates tumour cell invasion. Antibodies against serum amyloid A3 blocked metastasis.

## GEOLOGY

### Primitive petrous

*Science* **321**, 1828–1831 (2008)

A beige outcrop in northern Quebec may be Earth's oldest known crustal rock. Jonathan O'Neil of McGill University in Montreal, Canada, and his colleagues have dated parts of the stone using ratios of neodymium and samarium isotopes, and calculated the oldest section to be 4.28 billion years old. This is 250 million years older than the previous record-holder.

The rocks in question are from the Nuvvuagittuq greenstone belt. This belt had been estimated to be 3.8 billion years old, based on an analysis of zircon crystals. But the stone that O'Neil and his team probed contained no zircons, forcing them to use an alternative method. The outcrop's low levels of neodymium suggest that it formed before Earth's neodymium levels became fixed 4.1 billion years ago.

## MECHANICS

### Slippery when clean

*Phys. Rev. Lett.* **101**, 125505 (2008)

Friction is a familiar force in everyday life, but its nanometre-scale details are obscure. This is because the fundamental mechanisms are subtle and sensitive to contamination,



DINODIA IMAGES/ALAMY

say André Schirmeisen of the University of Münster, Germany, and his colleagues.

They pushed islands of the element antimony across a graphite surface using the tip of an atomic-force microscope. Some of the particles encountered frictional resistance proportional to their area of contact with the surface; others slid almost friction-free.

The latter state, called superlubricity, has been argued to arise from a mismatch between the atomic-scale corrugations of two surfaces, which, in theory, should be the norm for solids. Schirmeisen and his team conclude that lubricity is undermined by impurities stuck at the interface.

## PLANETARY SCIENCE

### Mars lander

*Icarus* **197**, 452–457 (2008)

A curious elongated crater in the northern lowlands of Mars may mark the final resting place of a lost moonlet. A related crater a short distance away and ‘butterfly wings’ of ejecta to either side show that the crater was formed by the larger of two objects following the same, shallow trajectory.

According to modelling by John Chappelow and Robert Herrick at the University of Alaska Fairbanks, the distance to the secondary crater makes it improbable that this was the impact of an asteroid that split up in the atmosphere. And the alignment of the crater and its secondary makes it unlikely to have been a double asteroid. A small moon brought down by tidal drag and fractured in the atmosphere is, they argue, the most likely source.

## ATMOSPHERIC CHEMISTRY

### A chemical equator

*J. Geophys. Res.* doi:10.1029/2008JD009940 (2008)

A narrow atmospheric boundary in the Western Pacific keeps apart the more polluted air of the Northern Hemisphere from the cleaner air of the south. This newfound divide is markedly farther north than the Intertropical Convergence Zone (ITCZ), a tropical low-pressure belt that is thought to separate air masses elsewhere according to their hemispheric origin.

Jacqueline Hamilton of the University of York, UK, and her team found that carbon

monoxide pollution from biomass burning in Thailand and Indonesia dropped steeply across the 50-kilometre-wide boundary.

They conclude that storms may lift air from the Northern Hemisphere into the upper troposphere — where pollutants remain longer — preventing it from mixing with southern air masses.

## THEORETICAL PHYSICS

### Computing with rainbows

*Phys. Res. Lett.* **101**, 130501 (2008)

Schemes for quantum computing abound, but most intend to carry out computations on objects such as atoms. Now Nicolas Menicucci at Princeton University in New Jersey and his colleagues propose a method that uses a rainbow of colours.

The group suggests firing lasers of 15 different frequencies into a cavity with a mirror at each end. Inside the cavity, a crystal splits each laser's photons into quantum mechanically ‘entangled’ pairs. Those pairs, in turn, become entangled with photons from the other lasers.

The resulting cobweb of entangled photons could be visualized as a brightly coloured tube (pictured left).

The authors would be able to manipulate their rainbow computer by measuring the entangled photons that escape from the cavity — and the computer could, in theory, perform any computation.

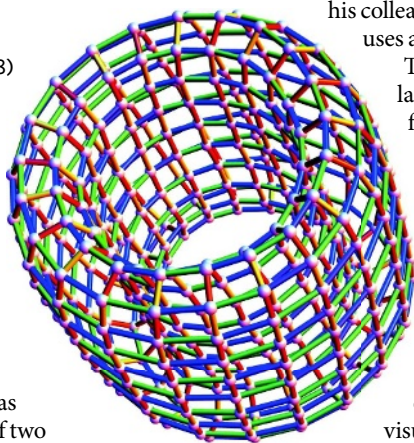
## CHEMISTRY

### Biofuel acid test

*Angew. Chem. Int. Edn* doi:10.1002/anie.200802879 (2008)

Tough, chewy parts of plants and even wood can be tapped for their fuel by dissolving them in an ionic liquid and then passing them over a solid acid catalyst, report Ferdi Schüth and his co-workers at the Max Planck Institute for Coal Research in Mülheim an der Ruhr, Germany. Specifically, a liquid made of an alkylmethylimidazolium salt dissolves woodchips. This allows the cellulose to be selectively hydrolysed when it passes through pores of a resin that contains sulphonic groups, generating sugars and smaller cellulose fragments.

The acidic resins needed to break down the cellulose are already commercially available, making the process easy to apply on a large scale.



N. C. MENICUCCI ET AL./PHYS. REV. LETT.

## JOURNAL CLUB

Roger Buick

University of Washington, Seattle

**An astrobiologist considers the implications of microbes' mining abilities.**

Microbes have been boring ever since life began on Earth: boring into rocks, that is. But why? Perhaps to avoid competitors or predators, to escape from environmental extremes or simply to secure a site safe from turbulent waters. Or might they be mining minerals for essential nutrients? Although the reason may vary depending on environment and host rock, a recent paper shows that some microbes tunnel towards a particular mineral, suggesting that nutrient mining may be occurring.

Tony Walton of the University of Kansas in Lawrence describes (and illustrates, gloriously) microscopic tubes in submarine glassy basalts from Hawai'i that show all the complex features of microbial borings (A. W. Walton *Geobiology* **6**, 351–364; 2008). The boreholes converge on olivine microcrystals but avoid plagioclase like the plague.

Olivine incorporates trace metals such as nickel, copper and chromium, essential nutrients for many microbes because they form the reactive centres in metalloenzymes and cofactors that catalyse key steps in vital metabolic pathways. These metals are sensitive to levels of oxygen and sulphides, so their bioavailability may have changed as Earth's surface environment has become more oxygenated and, periodically, more or less sulphidic. So the microbes may be mining olivine for metals that are now or were once rare in solution.

Two implications arise. First, although hominids have shown an ability to recognize different rocks for almost a million years, this geological aptitude may be more widespread and more ancient among other organisms. And second, as olivine occurs in martian meteorites and on Mars' surface, perhaps future astrobiological space missions should be alert to the possibility that fossils of microbial miners may occur in subaqueously deposited basaltic sands on that planet.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>



say André Schirmeisen of the University of Münster, Germany, and his colleagues.

They pushed islands of the element antimony across a graphite surface using the tip of an atomic-force microscope. Some of the particles encountered frictional resistance proportional to their area of contact with the surface; others slid almost friction-free.

The latter state, called superlubricity, has been argued to arise from a mismatch between the atomic-scale corrugations of two surfaces, which, in theory, should be the norm for solids. Schirmeisen and his team conclude that lubricity is undermined by impurities stuck at the interface.

## PLANETARY SCIENCE

### Mars lander

*Icarus* **197**, 452–457 (2008)

A curious elongated crater in the northern lowlands of Mars may mark the final resting place of a lost moonlet. A related crater a short distance away and ‘butterfly wings’ of ejecta to either side show that the crater was formed by the larger of two objects following the same, shallow trajectory.

According to modelling by John Chappelow and Robert Herrick at the University of Alaska Fairbanks, the distance to the secondary crater makes it improbable that this was the impact of an asteroid that split up in the atmosphere. And the alignment of the crater and its secondary makes it unlikely to have been a double asteroid. A small moon brought down by tidal drag and fractured in the atmosphere is, they argue, the most likely source.

## ATMOSPHERIC CHEMISTRY

### A chemical equator

*J. Geophys. Res.* doi:10.1029/2008JD009940 (2008)

A narrow atmospheric boundary in the Western Pacific keeps apart the more polluted air of the Northern Hemisphere from the cleaner air of the south. This newfound divide is markedly farther north than the Intertropical Convergence Zone (ITCZ), a tropical low-pressure belt that is thought to separate air masses elsewhere according to their hemispheric origin.

Jacqueline Hamilton of the University of York, UK, and her team found that carbon

monoxide pollution from biomass burning in Thailand and Indonesia dropped steeply across the 50-kilometre-wide boundary.

They conclude that storms may lift air from the Northern Hemisphere into the upper troposphere — where pollutants remain longer — preventing it from mixing with southern air masses.

## THEORETICAL PHYSICS

### Computing with rainbows

*Phys. Res. Lett.* **101**, 130501 (2008)

Schemes for quantum computing abound, but most intend to carry out computations on objects such as atoms. Now Nicolas Menicucci at Princeton University in New Jersey and his colleagues propose a method that uses a rainbow of colours.

The group suggests firing lasers of 15 different frequencies into a cavity with a mirror at each end. Inside the cavity, a crystal splits each laser's photons into quantum mechanically ‘entangled’ pairs. Those pairs, in turn, become entangled with photons from the other lasers.

The resulting cobweb of entangled photons could be visualized as a brightly coloured tube (pictured left).

The authors would be able to manipulate their rainbow computer by measuring the entangled photons that escape from the cavity — and the computer could, in theory, perform any computation.

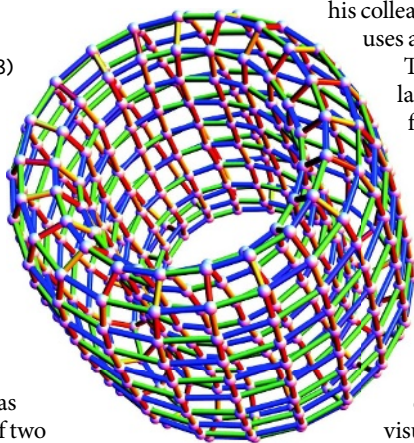
## CHEMISTRY

### Biofuel acid test

*Angew. Chem. Int. Edn* doi:10.1002/anie.200802879 (2008)

Tough, chewy parts of plants and even wood can be tapped for their fuel by dissolving them in an ionic liquid and then passing them over a solid acid catalyst, report Ferdi Schüth and his co-workers at the Max Planck Institute for Coal Research in Mülheim an der Ruhr, Germany. Specifically, a liquid made of an alkylmethylimidazolium salt dissolves woodchips. This allows the cellulose to be selectively hydrolysed when it passes through pores of a resin that contains sulphonic groups, generating sugars and smaller cellulose fragments.

The acidic resins needed to break down the cellulose are already commercially available, making the process easy to apply on a large scale.



N. C. MENICUCCI ET AL./PHYS. REV. LETT.

## JOURNAL CLUB

Roger Buick

University of Washington, Seattle

### An astrobiologist considers the implications of microbes' mining abilities.

Microbes have been boring ever since life began on Earth: boring into rocks, that is. But why? Perhaps to avoid competitors or predators, to escape from environmental extremes or simply to secure a site safe from turbulent waters. Or might they be mining minerals for essential nutrients? Although the reason may vary depending on environment and host rock, a recent paper shows that some microbes tunnel towards a particular mineral, suggesting that nutrient mining may be occurring.

Tony Walton of the University of Kansas in Lawrence describes (and illustrates, gloriously) microscopic tubes in submarine glassy basalts from Hawai'i that show all the complex features of microbial borings (A. W. Walton *Geobiology* **6**, 351–364; 2008). The boreholes converge on olivine microcrystals but avoid plagioclase like the plague.

Olivine incorporates trace metals such as nickel, copper and chromium, essential nutrients for many microbes because they form the reactive centres in metalloenzymes and cofactors that catalyse key steps in vital metabolic pathways. These metals are sensitive to levels of oxygen and sulphides, so their bioavailability may have changed as Earth's surface environment has become more oxygenated and, periodically, more or less sulphidic. So the microbes may be mining olivine for metals that are now or were once rare in solution.

Two implications arise. First, although hominids have shown an ability to recognize different rocks for almost a million years, this geological aptitude may be more widespread and more ancient among other organisms. And second, as olivine occurs in martian meteorites and on Mars' surface, perhaps future astrobiological space missions should be alert to the possibility that fossils of microbial miners may occur in subaqueously deposited basaltic sands on that planet.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

## NEWS

# NIH soon to be leaderless

Plaudits for departing director Elias Zerhouni may be echoing through the US National Institutes of Health (NIH) in Bethesda, Maryland — but underlying them is uncertainty about who will take over, and when. The White House has not yet named an acting director.

After six and a half years at the helm of the NIH, the world's largest biomedical research agency, Zerhouni announced last week that he will leave by the end of October. With this announcement, he sidestepped any notion that his decision is linked to the outcome of the 4 November presidential election. But it ushers in a transition period that will stretch for at least several months. The next US president will not take office until 20 January 2009, and high-level presidential nominations like that of NIH director can be achingly slow to make.

"We are all worried about what is going to happen in the interim and who the next director of NIH will be," says Story Landis, director of the National Institute of Neurological Disorders and Stroke, one of the NIH's 27 institutes and centres.

Zerhouni leaves as the \$29-billion agency faces great financial stress. Its budget doubled between 1998 and 2003, but since then its purchasing power has eroded by 10% as slight budget increases have failed to keep up with biomedical inflation. Many say that Zerhouni's work in the face of nearly flat funding has, of necessity, been the defining feature of his directorship (see 'Difficult times to make an impact').

Anthony Fauci, the long-time director of the National Institute of Allergy and Infectious Diseases, remembers advising the newly appointed



Elias Zerhouni: leaving this month.

Zerhouni: "Elias, what happens to you is going to rely very heavily on circumstances that are totally beyond your control." The two men still joke about it.

Zerhouni had faced challenges before. As the fifth of seven sons of a homemaker and a maths and physics teacher, he arrived in the United States from his native Algeria at the age of 24 with \$369 in his pocket. By the time he was recruited to the NIH in 2002, he was one of the top experts in magnetic resonance imaging (MRI), and, among other things, had pioneered

magnetic tagging — an MRI method that can be used to track heart motions in three dimensions. He had also risen to become executive vice-dean and chair of radiology at the Johns Hopkins School of Medicine in Baltimore, Maryland.

But it was soon apparent that the NIH gig wouldn't be a cake walk. "He comes into NIH and almost as soon as he gets there the good old days are over," says Howard Garrison, public-affairs director at the Federation of American Societies for Experimental Biology (FASEB) in Bethesda. As the agency's budget stagnated, success rates for grant applicants — especially first-time grant-seekers — plummeted. Zerhouni responded in 2006 with the 'Pathway to Independence' awards for young scientists, and managed to bring the number of first-time awards back up to 1,600 last year after it had dropped below 1,400 in 2006.

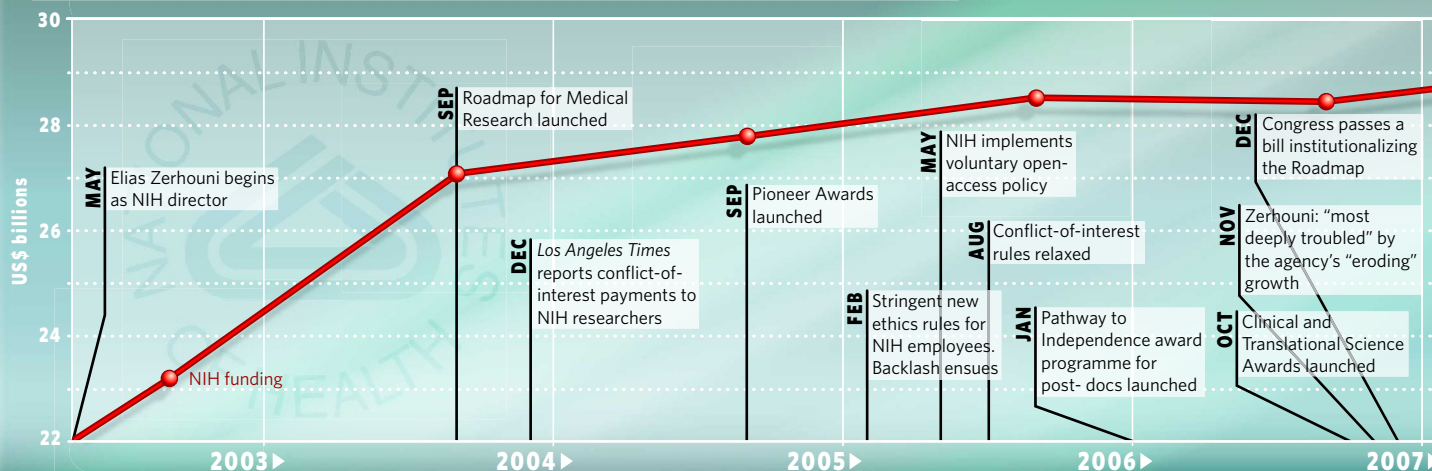
One early and much-criticized initiative was his 'Roadmap for Medical Research', a series of measures promoting trans-institute, high-risk, innovative research. As the budget for this grew from \$132 million in 2004 (0.47% of the total NIH budget) to \$495 million (1.7% of the total NIH budget) in 2008, it was perceived by some as too costly during a time of scarcity. In April 2006, Andrew Marks, then editor-in-chief of the *Journal of Clinical Investigation*, penned an angry editorial that began by telling Zerhouni: "Obviously you are not a scientist."

To this day, Zerhouni remains unfazed by criticism of the Roadmap. "I needed to do something to recognize that the boundaries of science have changed," he says.

Zerhouni also battled a conflict-of-interest

R. L. WOLLENBERG/NEWS.COM

## DIFFICULT TIMES TO MAKE AN IMPACT





**HAVE YOUR SAY**

Comment on any of our news stories, online.

[www.nature.com/news](http://www.nature.com/news)

scandal at the agency, after Congressional examiners uncovered lucrative payments to moonlighting intramural NIH researchers by drug companies with financial stakes in agency recommendations or research. Zerhouni implemented tough new ethics rules for staff scientists — which he softened only a little after an outcry on the Bethesda campus.

“He’s had to manage great expectations and stagnant resources,” says Tony Mazzaschi, senior director of scientific affairs at the Association of American Medical Colleges in Washington DC. “And in that environment, he was able to add power to the director’s role. That’s not any mean feat.”

The question is who will come next. Washington is rife with speculation, most of which will turn out to be wrong. “When Elias became director, his name was nowhere on anybody’s radar screen,” says Fauci. Fauci’s own name is inevitably floated whenever the NIH directorship is vacant. And any shortlist could include a stable of current and former institute heads, from heart institute chief Elizabeth Nabel to Francis Collins, who recently departed as director of the genome institute.

One other top medical spot in the Washington area got filled this week. Robert Tjian, a biochemist at the University of California, Berkeley, will take over the presidency of the Howard Hughes Medical Institute in Chevy Chase, Maryland, next spring from departing leader Tom Cech.

Zerhouni’s expected choice of acting director in his wake, deputy director Raynard Kington, is said to be a finalist for the chancellorship of the State University of New York. Doubtless the wisest shortlist reads: to be announced.

**Meredith Wadman**

See Editorial, page 565.

## Hwang work granted patent

Australia is to grant a patent for Woo Suk Hwang’s cloning method, even though the Korean scientist lied about using it to create human embryonic stem cells. But the patent is unlikely to prevent researchers from carrying out such work.

In 2004 and 2005, while at Seoul National University, South Korea, Hwang published a series of papers in which he claimed to have created a stem-cell line from a cloned embryo. An international patent describing his method was filed in 2004 by the university’s patent office. The application was based on an embryonic-stem-cell line that Hwang’s team had produced and deposited in an official stem-cell bank in accordance with the Budapest Treaty, which oversees the depositing of biological organisms for patent purposes.

In fact, the stem-cell line had been created not from a cloned embryo, but by a process called parthenogenesis, in which an egg develops into an embryo without being fertilized. Hwang was later charged with fraud, embezzlement and violation of the country’s bioethics laws, he was sacked from the university and his high-profile papers were editorially retracted because of their fabricated data. Proceedings against him are ongoing.

In June 2006, six months after Hwang’s work was discredited, the university’s patent office made applications in eleven countries, most of which were refused. But the patent passed all the requirements of Australia’s patent office, IP Australia: it was new, inventive, fully described and adequately defined.

IP Australia does not check for utility — that is, whether the patented procedure can actually produce what it claims. A representative there says there is no way they could test every claim that comes across their desks. Unlike most countries, the Australian patent office does not require authors to sign statements saying that their data are true.

IP Australia announced it was accepting the patent on 12 June, pending its standard 3-month period in which others can oppose it. No one opposed it.

Because of the extraordinary circumstances of this patent, it is now ‘on hold’. IP Australia has another 3 months to grant the patent. During that period, the applicant could withdraw or amend

it, or some “overriding right to refuse” could deny it. IP Australia is continuing to investigate the matter, but according to the representative, it is likely to be granted.

“There is no statutory basis to refuse to grant a patent on the basis that the scientific data in a patent application is a misrepresentation or fraudulently obtained,” wrote David Johnson, acting commissioner of patents at IP Australia, in a statement last week.

But Australia should refuse the patent on other grounds, according to David Earp, chief patent lawyer at Geron, the California company that holds international rights — including Australia — to an earlier patent that covers the cloning technique used to produce Dolly. “Geron retains all rights for use of [the cloning procedure] in human application, including the creation of embryonic stem cells,” he says.

“The broad claims of the recently accepted Hwang patent are not distinguishable from the [Dolly cloning] technology, and so the decision by the

Australian patent office to grant them appears to have been in error,” Earp says.

The patent is unlikely to be a powerful one. It would come into play only if the

university’s patent office tried to restrict a group in Australia from using the method. But such a group could challenge the patent in court on the grounds of utility, noting that the data were fraudulent and that the cell lines were derived from a parthenote, not a clone.

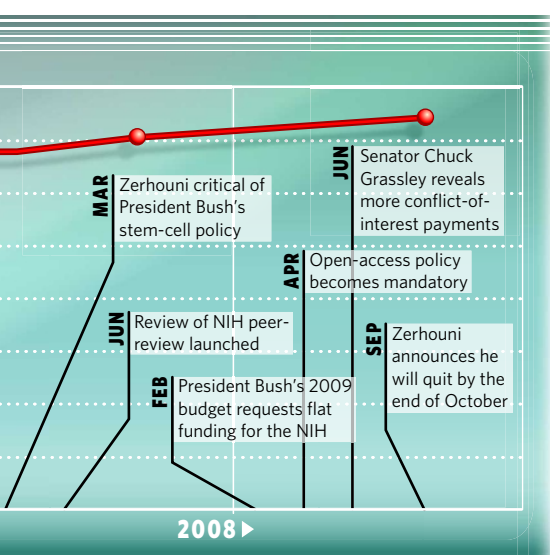
Johnson points out that even though misrepresentation cannot stop a patent from being granted, it “is grounds for revocation by the Court”. He adds that “IP Australia is not endorsing the research that underpins the application”.

The university’s patent office has applications pending in the United States, Canada, India and China.

Only a few people around the world are currently experimenting with human embryonic cloning. “Until a thorough investigation into the patent and its claims has been completed we cannot make any conclusions about the impact it would have on our project,” says Julia Schaft of the firm Sydney IVF, which last month became the first Australian group to receive a licence to attempt the technique.

**David Cyranoski**

**“IP Australia is not endorsing the research that underpins the patent.”**





**HAVE YOUR SAY**  
Comment on any of our  
news stories, online.  
[www.nature.com/news](http://www.nature.com/news)

scandal at the agency, after Congressional examiners uncovered lucrative payments to moonlighting intramural NIH researchers by drug companies with financial stakes in agency recommendations or research. Zerhouni implemented tough new ethics rules for staff scientists — which he softened only a little after an outcry on the Bethesda campus.

“He’s had to manage great expectations and stagnant resources,” says Tony Mazzaschi, senior director of scientific affairs at the Association of American Medical Colleges in Washington DC. “And in that environment, he was able to add power to the director’s role. That’s not any mean feat.”

The question is who will come next. Washington is rife with speculation, most of which will turn out to be wrong. “When Elias became director, his name was nowhere on anybody’s radar screen,” says Fauci. Fauci’s own name is inevitably floated whenever the NIH directorship is vacant. And any shortlist could include a stable of current and former institute heads, from heart institute chief Elizabeth Nabel to Francis Collins, who recently departed as director of the genome institute.

One other top medical spot in the Washington area got filled this week. Robert Tjian, a biochemist at the University of California, Berkeley, will take over the presidency of the Howard Hughes Medical Institute in Chevy Chase, Maryland, next spring from departing leader Tom Cech.

Zerhouni’s expected choice of acting director in his wake, deputy director Raynard Kington, is said to be a finalist for the chancellorship of the State University of New York. Doubtless the wisest shortlist reads: to be announced.

**Meredith Wadman**

See Editorial, page 565.

## Hwang work granted patent

Australia is to grant a patent for Woo Suk Hwang’s cloning method, even though the Korean scientist lied about using it to create human embryonic stem cells. But the patent is unlikely to prevent researchers from carrying out such work.

In 2004 and 2005, while at Seoul National University, South Korea, Hwang published a series of papers in which he claimed to have created a stem-cell line from a cloned embryo. An international patent describing his method was filed in 2004 by the university’s patent office. The application was based on an embryonic-stem-cell line that Hwang’s team had produced and deposited in an official stem-cell bank in accordance with the Budapest Treaty, which oversees the depositing of biological organisms for patent purposes.

In fact, the stem-cell line had been created not from a cloned embryo, but by a process called parthenogenesis, in which an egg develops into an embryo without being fertilized. Hwang was later charged with fraud, embezzlement and violation of the country’s bioethics laws, he was sacked from the university and his high-profile papers were editorially retracted because of their fabricated data. Proceedings against him are ongoing.

In June 2006, six months after Hwang’s work was discredited, the university’s patent office made applications in eleven countries, most of which were refused. But the patent passed all the requirements of Australia’s patent office, IP Australia: it was new, inventive, fully described and adequately defined.

IP Australia does not check for utility — that is, whether the patented procedure can actually produce what it claims. A representative there says there is no way they could test every claim that comes across their desks. Unlike most countries, the Australian patent office does not require authors to sign statements saying that their data are true.

IP Australia announced it was accepting the patent on 12 June, pending its standard 3-month period in which others can oppose it. No one opposed it.

Because of the extraordinary circumstances of this patent, it is now ‘on hold’. IP Australia has another 3 months to grant the patent. During that period, the applicant could withdraw or amend

it, or some “overriding right to refuse” could deny it. IP Australia is continuing to investigate the matter, but according to the representative, it is likely to be granted.

“There is no statutory basis to refuse to grant a patent on the basis that the scientific data in a patent application is a misrepresentation or fraudulently obtained,” wrote David Johnson, acting commissioner of patents at IP Australia, in a statement last week.

But Australia should refuse the patent on other grounds, according to David Earp, chief patent lawyer at Geron, the California company that holds international rights — including Australia — to an earlier patent that covers the cloning technique used to produce Dolly. “Geron retains all rights for use of [the cloning procedure] in human application, including the creation of embryonic stem cells,” he says.

“The broad claims of the recently accepted Hwang patent are not distinguishable from the [Dolly cloning] technology, and so the decision by the

Australian patent office to grant them appears to have been in error,” Earp says.

The patent is unlikely to be a powerful one. It would come into play only if the

university’s patent office tried to restrict a group in Australia from using the method. But such a group could challenge the patent in court on the grounds of utility, noting that the data were fraudulent and that the cell lines were derived from a parthenote, not a clone.

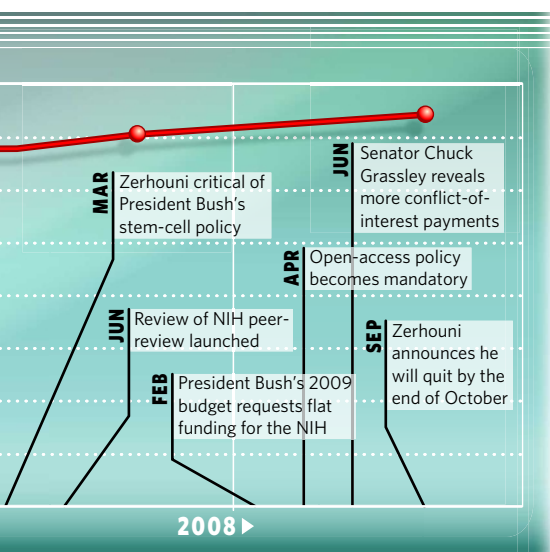
Johnson points out that even though misrepresentation cannot stop a patent from being granted, it “is grounds for revocation by the Court”. He adds that “IP Australia is not endorsing the research that underpins the application”.

The university’s patent office has applications pending in the United States, Canada, India and China.

Only a few people around the world are currently experimenting with human embryonic cloning. “Until a thorough investigation into the patent and its claims has been completed we cannot make any conclusions about the impact it would have on our project,” says Julia Schaft of the firm Sydney IVF, which last month became the first Australian group to receive a licence to attempt the technique.

**David Cyranoski**

**“IP Australia is not endorsing the research that underpins the patent.”**





# Fears surface over methane leaks

D. VAUGHAN/SCIENCE PHOTO LIBRARY

Preliminary data from two Arctic cruises suggest that rising temperatures are already causing substantial amounts of methane to be released from beneath the ocean floor. But catastrophic gas leaks, like those believed to have occurred 55 million years ago, are unlikely, scientists say.

In the past few weeks, scientists aboard the British research ship *James Clark Ross* have discovered more than 250 plumes of methane bubbling up along the continental margin northwest of Svalbard. The findings add to a similar discovery by a Russian team in August, that reported elevated methane concentrations near the Lena River delta, as part of the International Siberian Shelf Study (ISSS).

The findings have provoked alarmist media reports predicting massive methane bursts that could accelerate global warming. Methane is a far more powerful greenhouse gas than carbon dioxide, although it is present in much lower concentrations in the atmosphere.

But the phenomenon is probably not new. The scientists believe that methane has been released in the region for at least 15,000 years. "What we're now seeing certainly did not start in the last year or so," says geophysicist Graham Westbrook of the University of Birmingham, UK, who led the British team.

"We have observed increased methane concentrations in the Laptev Sea during several



British researchers found more than 250 plumes of methane bubbling up in the sea northwest of Svalbard.

expeditions since the mid-1990s," says Igor Semiletov, who oversees the ISSS methane programme aboard the Russian research ship *Jacob Smirnitskiy*. "But the data set is extremely limited. Whether what we're seeing

in the region is of any relevance for the global climate is mere speculation."

Semiletov says that the scientists did measure higher concentrations of dissolved methane this summer compared to summer

# Credit crunch threatens US wind-energy projects

Wind developers in the United States could be the first among the energy sector to fall victim to the global financial meltdown emanating from Wall Street.

The banking crisis that began with bad loans in the US housing sector has now brought down several commercial banks, one of the world's largest insurance companies and leading investment banks. These last have been particularly important in funding advanced energy technologies (through partnerships with wind developers) and in promoting international carbon markets.

The economic crisis could hit the booming US wind sector first, because the tax incentives designed to promote investment in the technology are meaningless for companies without sufficient profit. Current US law provides wind developers with a tax credit of 2 cents per kilowatt hour. That can add up to millions of dollars annually, but many wind operators do not



Wall Street's woes may halt the wind-energy boom.

have enough income to take full advantage of that type of tax credit.

Before declaring itself bankrupt on 15 September, US investment bank Lehman Brothers was one of several major firms that invested in wind projects in exchange for the tax credit, which they used to reduce their federal tax bill. The number of firms making such 'tax equity' investments has dropped from more

than a dozen to five or six in recent months, says Ethan Zindler, who heads up North American research for the London-based consultancy firm New Energy Finance. "To use the tax credit, you have to have tax exposure, and to have tax exposure you have to have profits."

Congress is fuelling anxiety; it has so far failed to extend the tax credit, which expires at the end of the year. Companies are on track to secure as much as \$8 billion–10 billion in tax-equity deals — up from \$5.2 billion in 2007 — assuming they can find the investors. The tax legislation would extend the credit for wind by one year and solar developers would receive an eight-year extension of a separate investment tax credit.

Most expect that the credit will ultimately be extended, even if it is allowed to lapse, but Zindler says that companies are nonetheless rushing to get their deals done now. Forays into carbon markets could become more difficult across the board as banking institutions curb their appetite

S. PLATT/GETTY

sampling in 2003 and 2004 (N. Shakhova and I. Semiletov *J. Mar. Sys.* **66**, 227–243; 2007). At one ice-covered site in the mere 50-metre shelf water, they detected methane bubbling at the surface, indicating that at least some of the gas released at the seabed is escaping into the atmosphere before being consumed by bacteria in the water column.

Geologists think that billions of tonnes of methane lie beneath the sub-sea permafrost in some parts of the shallow Siberian shelf, although estimates vary widely. The hydrocarbon — trapped there either as a gas, or bound in solid ice-like structures called methane hydrates — is a remnant from the last ice age when the sea level was about 100 metres lower. The big fear is that the methane could escape as a result of the permafrost becoming porous, possibly from an increased influx of freshwater from the relatively warm Lena River.

"The risk is real," says Hans-Wolfgang Hubberten, a permafrost expert at the Alfred Wegener Institute of Polar and Marine Research in Potsdam, Germany. "But there's no reason to panic. Claims that gas hydrates are on the brink of dissociating in a big way should be taken with a large pinch of salt."

Thermal modelling suggests that the marine permafrost in the region is relatively stable. However, drillings conducted in 2005 revealed that the permafrost may have slightly warmed and thinned (V. Rachold *et al. Eos* **88**, 149–156; 2007). Even so, says Hubberten, it is

likely that the observed emissions come from 'new' methane produced by increased bacterial activity in thawing soil, rather than from degradation of ancient gas hydrates.

Methane, air and water samples taken by both teams will now be sent to isotope labs in the Netherlands and the United Kingdom to help determine the source of the methane. Geochemical analysis should also show how much of the gas escapes to the atmosphere, says Westbrook. "The new findings will be useful in helping us assess the history of climate change in the region, and how the methane reservoirs responded to past temperature changes."

**"The risk is real, but there's no reason to panic."**

Globally, atmospheric methane concentrations increased by 7.5 parts per billion to nearly 1,800 parts per billion during 2007 after almost zero growth since 1999. The upward trend is likely to continue this year, says Ed Dlugokencky, who oversees the methane database run by the National Oceanographic and Atmospheric Administration (NOAA) in Boulder, Colorado. "Our data suggest increased emissions in the Arctic and the tropics," he says. "Both regions were apparently warmer and wetter than average."

Data collected by NOAA at remote sites are usually at least 6 weeks out of date. And NOAA's measurement network in the Arctic is not dense enough to tell if increased methane emissions come from wetlands, permafrost or from gas hydrates on the continental shelves. ■

Quirin Schiermeier

for risk. In particular, says Abyd Karmali, global head of carbon emissions at Merrill Lynch in London, the development of pilot projects using carbon trading to curb deforestation could get more difficult — at least until an international policy is put in place to guide such investments. Merrill Lynch, which is expected to merge with the Bank of America as a result of the ongoing crisis, is already developing one such project in Indonesia.

Harvard University economist Robert Stavins says that the financial crisis will probably have an impact on the voluntary carbon markets, especially if the economy dives into recession. But regulated markets, such as Europe's carbon trading scheme, will be fine, he says. "Compliance activity by business is immune to business cycles."

Some fear a prolonged economic crisis could



make it harder for the United States to enact global-warming legislation owing to concerns about even higher energy prices to come. David Victor, an energy policy expert at Stanford University in Palo Alto, California, says the danger is real: people who are short on cash are less willing to spend money on dealing with distant threats such as global warming.

But Victor believes the fundamental drivers behind the renewed interest in clean

energy — high oil prices and concerns about global warming — will remain. "We still have a whole bunch of renewable-energy technologies that are improving their performance," he says. "It's highly likely that people will find capital for these projects." ■

Jeff Tollefson

See Editorial, page 565.





## STING STORY

Jellyfish's chemical attack  
bequeathed by bacteria.  
[www.nature.com/news](http://www.nature.com/news)

NHPA/A.N.T. PHOTO LIBRARY

# Ancient water sites for next rover

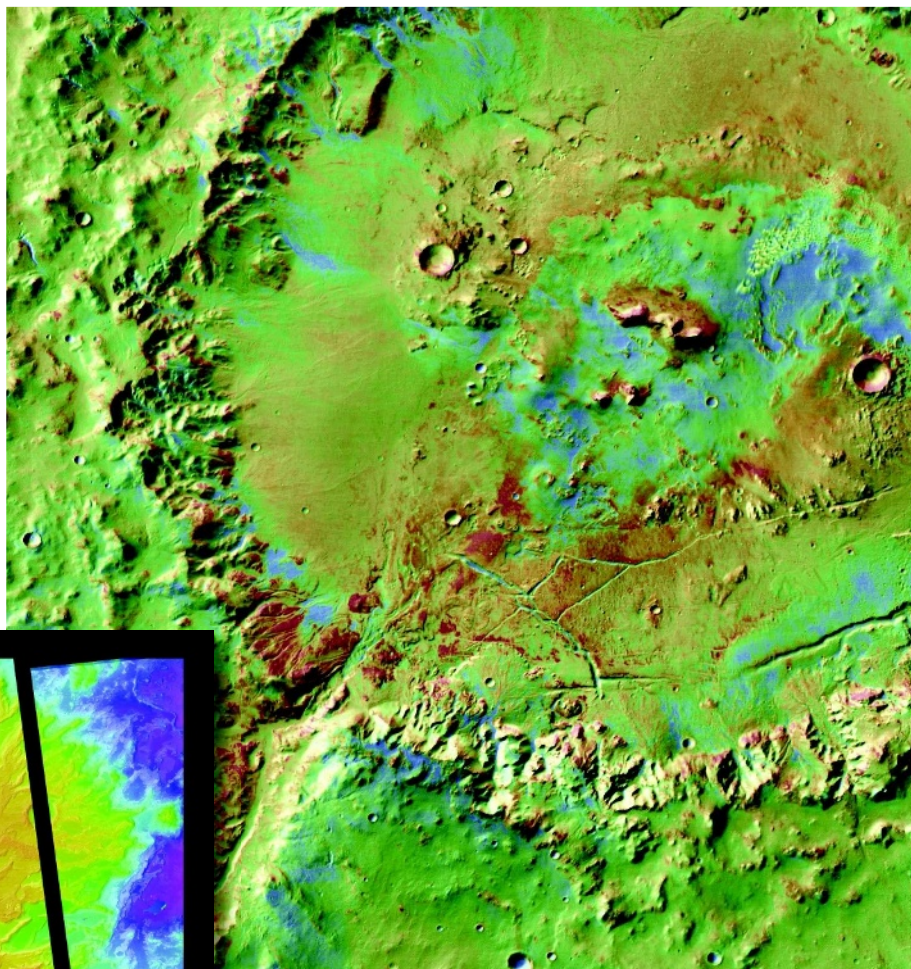
The next Mars rover could end up down in the delta if a group of geologists and astrobiologists get their way.

The Mars science community has ranked a landing site called Eberswalde Crater as the most tantalizing destination for the Mars Science Laboratory (MSL), the US\$2-billion rover that is due to launch in 2009. The crater seems to contain the remnants of a meandering river that spilled into a lake more than 3 billion years ago and piled up delta sediments — a prime target for MSL's instruments and their search for past or current microbial life. "If you go to any lake bed on Earth, that's where you find fossils," says James Rice, an astrogeologist at Arizona State University in Tempe who is a chief advocate for the site.

Scientists met last week to evaluate seven favoured sites (see 'Experts' shortlist') at a workshop in Monrovia, California. At the end of the meeting 104 paper ballots were cast, based only the scientific potential of each site — graded in 11 categories including the diversity of minerals likely to be present and the potential of the site to preserve evidence of life. At a meeting in November, the engineering team will rate the sites technically, comparing the risks of landing the 900-kilogram rover within 20-by-25-kilometre ellipses and driving it to the most interesting spots within a Martian year — nearly two Earth years — of operation.

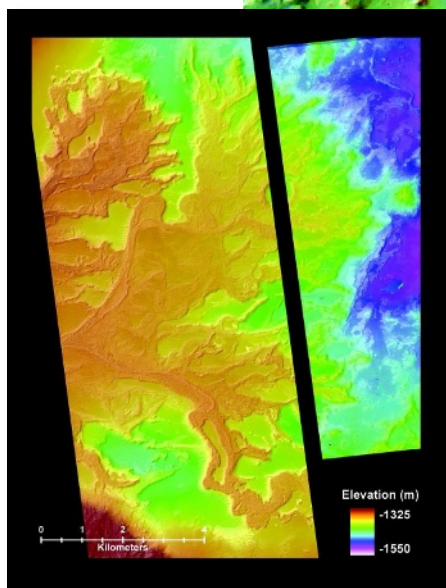
Engineers are worried about Eberswalde and the next most highly ranked site, Holden Crater, because they sit in more southerly latitudes of Mars. The rover will use a lubricant that might not work as well during winter temperatures at those latitudes.

But Rice and others are optimistic about the Eberswalde site's chances — it's made quite a comeback after almost being knocked out of consideration at a workshop last year. Rice says the site shows clear signs of ancient river channels, some 100 metres wide, that emptied into a lake 150 metres deep over a period of



K. LEWIS, CALTECH &amp; NASA/JPL/UNIV. ARIZONA

NASA/JPL/ARIZONA STATE UNIV.



The two top-ranked destinations for the next Mars rover are craters thought to contain ancient bodies of water. Eberswalde Crater, left, contains an ancient river delta, and the walls of Holden Crater, above, were breached by a flood that filled a lake in the centre.

perhaps a million years. The Holden Crater site lacks a delta but contains

eroded sections of a lake bed that the rover could access more quickly than the delta at Eberswalde, says John Grant, of the Smithsonian Institution in Washington DC, who is co-chair of the landing site steering committee. Grant, an advocate for Holden Crater, says its walls were later breached by a massive flood that added to the volume of the lake.

But one of the two top sites will probably face elimination at the next meeting because of the concerns over their similar southerly latitude, says Grant. After a workshop next April, the

committee will present recommendations to NASA science chief Edward Weiler, who will make the final decision on the target for the autumn 2009 launch.

**Eric Hand**

More images at <http://tinyurl.com/4vix98>.

## Experts' shortlist

- 1 Eberswalde Crater
- 2 Holden Crater
- 3 Gale Crater
- 4 Mawrth Vallis
- 5 Nili Fossae Trough
- 6 South Meridiani Planum
- 7 Miyamoto Crater



OBAMA v. MCCAIN

Read our election coverage online.

[www.nature.com/news](http://www.nature.com/news)

# Teams merge for dark-energy mission

A competition between groups hoping to design a space telescope to investigate how the Universe is expanding over time has been scrapped by NASA and the US Department of Energy (DoE).

Instead, the agencies are pursuing a government-built, government-led design for the Joint Dark Energy Mission (JDEM), which may accommodate elements from all three of the teams. "It's a do-over for all of us," says Michael Levi, who is co-principal investigator for the Supernova Acceleration Probe (SNAP), a team that, he says, all of a sudden doesn't really exist any more.

NASA had been giving money to SNAP and two other groups, called the Advanced Dark Energy Physics Telescope (ADEPT) and the Dark Energy Space Telescope (Destiny). Each team was pursuing a proprietary telescope design, emphasizing different methods (see 'The telescope teams') for seeking constraints on the mysterious energy that is thought to be accelerating the expansion of the Universe. The mission is pegged for launch in 2015.

However, on 12 September, NASA and the DoE announced they will develop a common "reference design" that would not preclude any of the three methods. The design will be worked out by a new programme office opened at Goddard Space Flight Center in Greenbelt, Maryland, and a science coordination group of 12–20 people. The membership could be



**Supernovae offer clues to the expanding Universe.**

decided by 3 October, according to NASA astrophysics division chief Jon Morse, who says that there were 50 applicants in total, some coming from all three teams. Neil Gehrels, principal investigator for the Swift Gamma Ray Burst Explorer at Goddard, will chair the group.

The decision took many by surprise. "I'm concerned," says Chuck Bennett, principal investigator for the ADEPT team at Johns Hopkins University in Baltimore, Maryland. "Three teams did a lot of work for a long time. I'm worried that hitting the reset button and starting again is going to set things back."

NASA has removed mention of the competition between the three projects from its websites; just weeks ago, it discussed

deciding between the telescopes in 2009.

It's not necessarily a bad move by the agencies, says Robert Cahn of Lawrence Berkeley National Laboratory in California, who was part of a task force convened in 2005 to examine the dark-energy question. The JDEM may have become too big and costly to have been managed well by the relatively small teams, but now NASA can adopt the best ideas from each, he says. "In some sense NASA seems to have made up its mind that it wants to do all three methods," he says. "It's certainly not working the way we expected but it might work out well."

Although the decision eliminates tension between the competing teams — all three presume they will share aspects of their once-secret designs for the science coordination group — there is still tension between what the scientists want to do, and how much NASA and the DoE say they can afford. A 2007 National Academies report that endorsed the JDEM estimated that the three designs would cost more than US\$1.3 billion in total. But Morse has said that he can afford only a \$600-million mission, not including launch costs. The DoE has said it wants to pay about 25% of the overall costs. Eleven of the academy report's authors complained to NASA and the DoE in May that the science they envisioned the JDEM doing would not be possible at half the cost. "Wishful thinking does not engineer successful spacecraft," they wrote.

**Eric Hand**

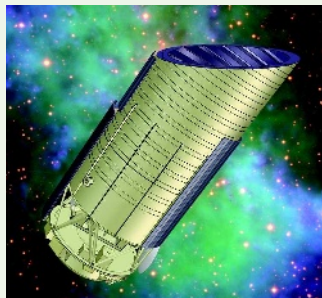
J.J. HESTER/ARIZONA STATE UNIV./NASA

## The telescope teams

Three main teams designed space telescopes to measure how the Universe's expansion rate has changed over time, by performing surveys of objects from early in its history. Surveys require wide fields of view, and the need to look far back in time means that the telescopes would have to see infrared, the light with which the most distant and early objects glow. Both tasks are difficult from Earth. The teams emphasized different targets and techniques:

### Supernovae

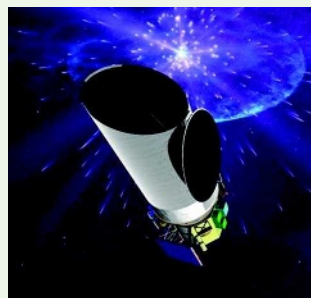
Dark energy was discovered in 1998 by using the assumption that all supernovae in a class shine with the same luminosity, which allows astronomers to calculate their distance from Earth more precisely. It was found



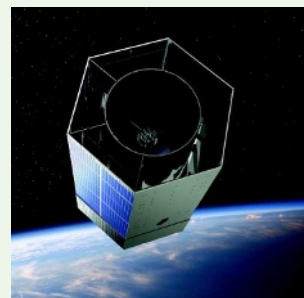
that more distant supernovae were receding more quickly than had been expected. The Supernova Acceleration Probe team initially emphasized this technique, although it later adopted all three methods.

### Weak gravitational lensing

A high-resolution picture of distant galaxies can reveal



tiny distortions in their shapes caused by the 'lensing' of intervening dark matter. The uneven distribution of this matter, and how it changed with cosmological time, would be a proxy for how dark energy influenced the Universe's growth. The Dark Energy Space Telescope team emphasized a blend of lensing and supernovae.

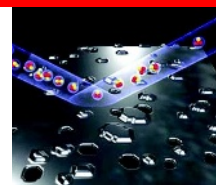


### Baryonic acoustic oscillations

Sound waves soon after the Big Bang created ripples in the distribution of galaxies. By comparing the ripples from galaxy clusters in the early Universe with ripples in later clusters, astronomers can deduce the effect of dark energy over time. The Advanced Dark Energy Physics Telescope team emphasized this approach. **E.H.**

NASA/GSFC



**ATOM REFLECTIONS**

Ultrasmooth mirror could herald birth of a new microscope.

[www.nature.com/news](http://www.nature.com/news)

D. BARREDO ET AL. ADV. MATER. 20, 3492–3497 (2008).

**SNAPSHOT****How do you like your coffee?**

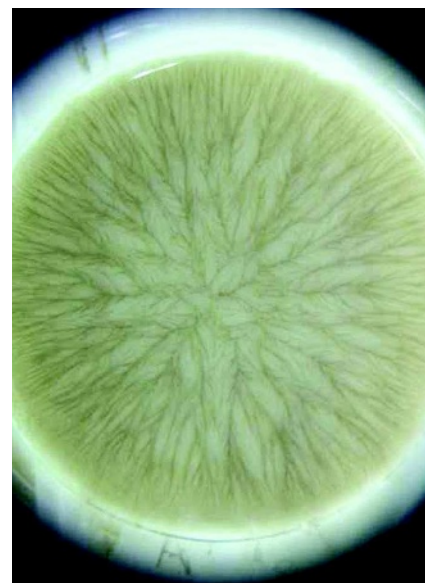
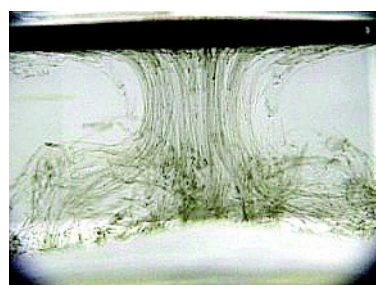
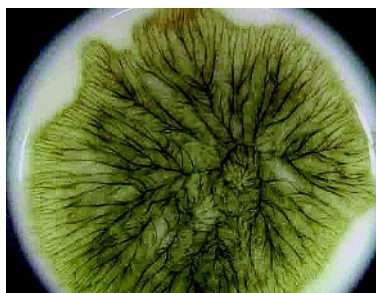
This floating fractal (top, left) is formed 90 seconds after a drop of instant coffee falls into a cup of milk.

Coffee is heavier than milk and the battle between gravity and surface tension plays out at the boundary between the two liquids. The coffee falls vertically through the milk (bottom, left, with water replacing milk for ease of viewing), and the fractal pattern emerges.

The pattern constantly shifts as parts of it are sucked into the milk, producing a fractal structure with the same dimension as a Sierpiński carpet — formed when a square is cut into nine identical squares; the central square is removed; and the procedure is repeated with the remaining eight squares and so on infinitely.

Michiko Shimokawa and Shonosuke Ohta, fluid scientists at Kyushu University in Fukuoka City, Japan, say that it is the first time this kind of fractal has been shown experimentally ([www.arxiv.org/abs/0809.2458](http://www.arxiv.org/abs/0809.2458)), and they managed to recreate the process using a magnetic liquid instead of coffee (far right).

Katharine Sanderson



## World nuclear security gets welcome boost

An international initiative to help safeguard nuclear materials worldwide was announced on 30 September. The World Institute for Nuclear Security plans to do for nuclear security what the World Association of Nuclear Operators, created in the aftermath of the 1986 Chernobyl power-plant accident, does for nuclear safety.

The institute will bring together nuclear players, including scientific experts, to share sensitive information. Its goal is to strengthen accounting, control and physical protection of nuclear materials and facilities worldwide. The group, to be based

in Vienna, is the brainchild of the Nuclear Threat Initiative group in Washington DC, co-chaired by former senator Sam Nunn and broadcast mogul Ted Turner. Roger Howsley, former director for security at British Nuclear Fuels, will be executive director.

## South Africa replaces its health minister

African National Congress member Barbara Hogan took over last week as South Africa's health minister. She replaces the controversial Manto Tshabalala-Msimang, who had sparked international outrage by proposing that HIV be treated with garlic and beetroot.

AIDS activists welcomed the appointment, made by the government of new president Kgalema Motlanthe. Hogan is on the advisory board of the Amandla AIDS Fund, which provides grants for HIV/AIDS prevention and treatment programmes.

But the South African parliament also quietly passed a law that gives the minister sweeping authority over the approval of new medicines and a remit to regulate traditional medicines alongside conventional pharmaceuticals. The bill created a body — the South African Health Products Regulatory Authority — to

oversee the approval of medicines.

Crucially, the agency's chief executive, who will be appointed by the health minister, will be accountable not to a board, as the existing Medicines Control Council is, but only to the minister. Critics of the bill fear that it may lead to conflicts of interest in the public-health system, and that it risks diluting the scientific basis behind making new treatments available.

For a longer version of this story, see <http://tinyurl.com/4hh928>.

## Falcon rocket reaches low-Earth orbit

It was fourth time lucky for Space Exploration Technologies and its Falcon 1 rocket. After a trio of failed flights since 2006, the privately funded rocket soared into low-Earth orbit from Kwajalein atoll on Omelek Island on 28 September.

SpaceX and its founder, Elon Musk, hope that the launch marks the dawn of a new, substantially cheaper era of space flight. The company estimates that each Falcon-1 launch will cost less than US\$10 million; existing systems can cost up to four times as much.

Next up are the company's larger heavy-lift rockets. Dubbed Falcon 9 and Falcon 9 Heavy, these rockets could potentially carry



Better safeguards are planned for nuclear material.

M. GHASEMI, ISNA/AP

## World nuclear security gets welcome boost

An international initiative to help safeguard nuclear materials worldwide was announced on 30 September. The World Institute for Nuclear Security plans to do for nuclear security what the World Association of Nuclear Operators, created in the aftermath of the 1986 Chernobyl power-plant accident, does for nuclear safety.

The institute will bring together nuclear players, including scientific experts, to share sensitive information. Its goal is to strengthen accounting, control and physical protection of nuclear materials and facilities worldwide. The group, to be based

in Vienna, is the brainchild of the Nuclear Threat Initiative group in Washington DC, co-chaired by former senator Sam Nunn and broadcast mogul Ted Turner. Roger Howsley, former director for security at British Nuclear Fuels, will be executive director.

## South Africa replaces its health minister

African National Congress member Barbara Hogan took over last week as South Africa's health minister. She replaces the controversial Manto Tshabalala-Msimang, who had sparked international outrage by proposing that HIV be treated with garlic and beetroot.

AIDS activists welcomed the appointment, made by the government of new president Kgalema Motlanthe. Hogan is on the advisory board of the Amandla AIDS Fund, which provides grants for HIV/AIDS prevention and treatment programmes.

But the South African parliament also quietly passed a law that gives the minister sweeping authority over the approval of new medicines and a remit to regulate traditional medicines alongside conventional pharmaceuticals. The bill created a body — the South African Health Products Regulatory Authority — to

oversee the approval of medicines.

Crucially, the agency's chief executive, who will be appointed by the health minister, will be accountable not to a board, as the existing Medicines Control Council is, but only to the minister. Critics of the bill fear that it may lead to conflicts of interest in the public-health system, and that it risks diluting the scientific basis behind making new treatments available.

For a longer version of this story, see <http://tinyurl.com/4hh928>.

## Falcon rocket reaches low-Earth orbit

It was fourth time lucky for Space Exploration Technologies and its Falcon 1 rocket. After a trio of failed flights since 2006, the privately funded rocket soared into low-Earth orbit from Kwajalein atoll on Omelek Island on 28 September.

SpaceX and its founder, Elon Musk, hope that the launch marks the dawn of a new, substantially cheaper era of space flight. The company estimates that each Falcon-1 launch will cost less than US\$10 million; existing systems can cost up to four times as much.

Next up are the company's larger heavy-lift rockets. Dubbed Falcon 9 and Falcon 9 Heavy, these rockets could potentially carry



Better safeguards are planned for nuclear material.

M. GHASEMI, ISNA/AP

much bigger payloads farther from Earth.

The first launch for Falcon 9, whose nine engines have been successfully fired on the ground, is slated for next year.

For a longer version of this story, see <http://tinyurl.com/43gnya>.

## Carbon dioxide emissions rise to record levels

Carbon dioxide emissions from fossil fuels and cement manufacturing are rising faster than the worst-case scenario drawn up by the Intergovernmental Panel on Climate Change (IPCC). According to the latest worldwide carbon budget, released by the Global Carbon Project, CO<sub>2</sub> levels rose by 3.5% a year between 2000 and 2007, compared with 2.7% as calculated by the IPCC. During the 1990s, emissions rose at 0.9% a year.

“For a decade we’ve been using the [IPCC] middle-ground scenario, while we’re actually in a different realm of emissions,” says Pep Canadell, the project’s executive director.

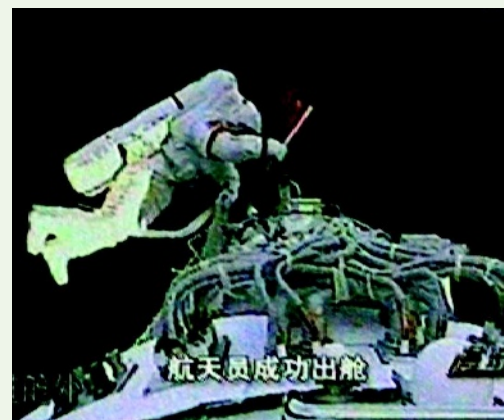
China is now the biggest emitter of CO<sub>2</sub> and responsible for 21% of the world’s emissions — up from 14% in 2002. This knocks the United States into second place, contributing 19% of global emissions. India is fourth, but looks set to take third place from Russia this year.

## China’s first spacewalk

Zhai Zhigang (pictured), the commander of the three-person Shenzhou VII spacecraft, has become the first Chinese astronaut, or ‘taikonaut’, to spacewalk. Zhai spent 13 minutes outside the orbital module on 27 September.

A fire alarm went off while he was conducting the spacewalk, but it turned out to be due to a faulty sensor. Zhai retrieved a rack containing lubricant samples from outside the spacecraft.

The mission landed in Mongolia on 28 September.



REUTERS

## US Congress approves funding bill for science

Most US science agencies will see their budgets frozen at 2008 levels under a massive \$630-billion spending bill that was passed by Congress on 27 September. The ‘continuing resolution’ keeps the government operating until March 2009, when a new president and new Congress will tackle funding priorities.

A few agencies did get new dollars for 2009. Science and technology funding within the Department of Defense, for instance, rose 7%, and research and development funding at the

Department of Homeland Security rose 9%.

The bill also provides NASA with permission to buy flights on Soyuz spacecraft from Russia until 2016, in order to ferry astronauts and cargo to the International Space Station after the space shuttle is retired.

### Corrections

In the News Feature ‘The fate of fingers’ (*Nature* 455, 160–164; 2008), we misspelled the name of Matthew Porteus of the University of Texas: apologies. In the News Feature ‘Postmodern evolution?’ (*Nature* 455, 281–284; 2008), we should have identified Susan Mazur as the first to use the term ‘The Woodstock of Evolution’ to describe the Altenberg meeting.



much bigger payloads farther from Earth.

The first launch for Falcon 9, whose nine engines have been successfully fired on the ground, is slated for next year.

For a longer version of this story, see <http://tinyurl.com/43gnya>.

## Carbon dioxide emissions rise to record levels

Carbon dioxide emissions from fossil fuels and cement manufacturing are rising faster than the worst-case scenario drawn up by the Intergovernmental Panel on Climate Change (IPCC). According to the latest worldwide carbon budget, released by the Global Carbon Project, CO<sub>2</sub> levels rose by 3.5% a year between 2000 and 2007, compared with 2.7% as calculated by the IPCC. During the 1990s, emissions rose at 0.9% a year.

“For a decade we’ve been using the [IPCC] middle-ground scenario, while we’re actually in a different realm of emissions,” says Pep Canadell, the project’s executive director.

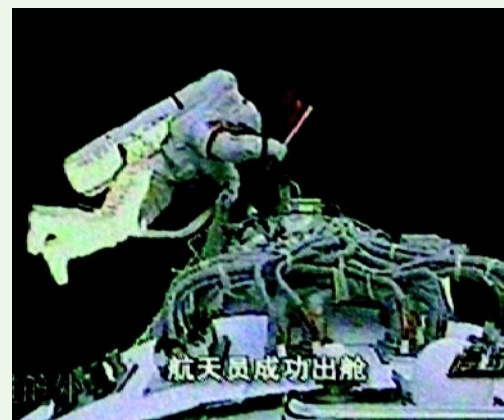
China is now the biggest emitter of CO<sub>2</sub> and responsible for 21% of the world’s emissions — up from 14% in 2002. This knocks the United States into second place, contributing 19% of global emissions. India is fourth, but looks set to take third place from Russia this year.

## China’s first spacewalk

Zhai Zhigang (pictured), the commander of the three-person Shenzhou VII spacecraft, has become the first Chinese astronaut, or ‘taikonaut’, to spacewalk. Zhai spent 13 minutes outside the orbital module on 27 September.

A fire alarm went off while he was conducting the spacewalk, but it turned out to be due to a faulty sensor. Zhai retrieved a rack containing lubricant samples from outside the spacecraft.

The mission landed in Mongolia on 28 September.



REUTERS

## US Congress approves funding bill for science

Most US science agencies will see their budgets frozen at 2008 levels under a massive \$630-billion spending bill that was passed by Congress on 27 September. The ‘continuing resolution’ keeps the government operating until March 2009, when a new president and new Congress will tackle funding priorities.

A few agencies did get new dollars for 2009. Science and technology funding within the Department of Defense, for instance, rose 7%, and research and development funding at the

Department of Homeland Security rose 9%.

The bill also provides NASA with permission to buy flights on Soyuz spacecraft from Russia until 2016, in order to ferry astronauts and cargo to the International Space Station after the space shuttle is retired.

### Corrections

In the News Feature ‘The fate of fingers’ (*Nature* 455, 160–164; 2008), we misspelled the name of Matthew Porteus of the University of Texas: apologies. In the News Feature ‘Postmodern evolution?’ (*Nature* 455, 281–284; 2008), we should have identified Susan Mazur as the first to use the term ‘The Woodstock of Evolution’ to describe the Altenberg meeting.



# The Pentagon's culture wars

What began several years ago as an attempt to recruit social scientists to help the military has sparked a broader debate about militarizing academia. **Sharon Weinberger** reports.

It is a story that repeats with grim monotony: US forces in Iraq detain a suspected insurgent after they find in his home what they think is jihadist literature and an illegal weapon. These detentions — often based on mistaken assumptions or poor intelligence — can easily escalate into major conflicts with the local community.

But in one recent case, researchers helped defuse a potential conflict. Analysts working for a 'human-terrain team' informed a US commander that the 'jihadist' literature discovered in the village of Banat al Hassan, about 30 kilometres northwest of Baghdad, was ordinary religious teaching material, and the weapon — a riflescope — was for a pellet gun that beekeepers in the area use for shooting birds. The suspect was promptly released, and his family ended up helping US forces by revealing the location of a large improvised explosive device.

This upbeat anecdote is "a story about how a little respect, culture and compassion can save human life", says Montgomery McFate,

an anthropologist at the Institute for Defense Analyses in Alexandria, Virginia, and senior adviser to the Pentagon's human-terrain programme. But it also underscores some of the complexities and controversies surrounding the Pentagon's quest for 'cultural knowledge'. What if, for example, the literature had indeed been jihadist literature? Would the human-terrain teams, which include civilian social scientists, then be helping the military to target insurgents?

Last year, the Pentagon provided almost \$60 million for the Human Terrain System, a Department of Defense programme that represents the latest incarnation of the military's long, troubled relationship with social science (see 'Lessons from the past', overleaf). It includes deployed teams that directly advise military commanders in the field, specialized software for cultural 'mapping' plus personnel based in the United States conducting research. According to official figures provided by the army, there are now sixteen five-person Human Terrain Teams (HTTs) deployed in Iraq

and five in Afghanistan, along with about 40 people in 'research reachback cells' in the United States. The teams are supposed to provide deployed military forces with "direct social-science support in the form of ethnographic and social research, cultural information research, and social data analysis".

But the effort is not without its problems; two social scientists have been killed in the field, one in Afghanistan and one in Iraq. And critics fear that this sort of work poses ethical problems, particularly if it's telling the military who is, or isn't, a potential enemy. Last November, the executive board of the American Anthropological Association (AAA) condemned the effort, saying it "creates conditions which are likely to place anthropologists in positions in which their work will be in violation of the AAA code of ethics", as well as endanger other anthropologists by bringing suspicion on their activities. The association is also proposing changes to its rules of ethics that would tighten restrictions on secret research.

Beyond the AAA, a number of researchers

F. MONTEFORTE/AFP/GETTY IMAGES



in 2007 founded the Network of Concerned Anthropologists, which asks colleagues to sign a pledge committing them to “refrain from directly assisting the US military in combat, be it through torture, interrogation, or tactical advice”. Though there hasn’t been any known case of that happening with the HTTs, historical precedents exist. During the Second World War, for instance, anthropologists helped raise guerilla armies, passed information used to plan bombing raids and theorized about race-specific bioweapons.

Critics say the current work flies in the face of everything anthropology represents, from transparency of research to informed consent (for example, the social scientists on the HTTs do not submit their research to an institutional review board, as would be normally required for human research). “I don’t think there’s a place for embedded anthropologists with combat missions,” says Roberto Gonzalez, an anthropologist at the University of California, Berkeley, who is working on a book about the Human Terrain System. “It runs completely counter to anthropology’s ethical framework, something that’s come about over a long, bitter period that goes back to the First World War.”

### Militarizing anthropology?

McFate has emerged as the most public face of the Pentagon’s military anthropology work. She got her PhD in anthropology from Yale University, focusing on the British counter-insurgency in Northern Ireland, and by 2005 she had co-authored an article in a military journal outlining a plan for deploying social science advisers with troops (M. McFate and A. Jackson *Military Rev.* July/Aug, 18–21; 2005). For her, the issue is unabashedly about moving anthropology toward an applied discipline that can aid the military. “Why should anthropology be some leftist religion?” she asks. “I mean, it’s supposed to be a science; it’s not supposed to be a political platform, a substitute for the Peace Corps, or a cult.”

The Pentagon, however, has had a hard time recruiting and keeping qualified anthropologists. Of 35 social scientists based in Iraq and Afghanistan, only about half have PhDs, and only seven of those deployed are anthropologists. One social scientist hired to work on a HTT was identified during screening as a convicted criminal (and dismissed prior



**“Why should anthropology be some leftist religion? I mean, it’s supposed to be a science.”**  
— Montgomery McFate

to deployment), another was found medically unfit, two were let go because of security clearance issues, and two were fired for performance issues. The company responsible for hiring the researchers is BAE Systems, a major Pentagon contractor, and some have criticized its focus on recruiting through intelligence and military-focused websites, as opposed to academic venues.

One of those fired was Zenia Helbig, a PhD candidate in religious studies, who says she was let go by BAE after a joking comment she made over drinks with colleagues about switching sides if the United States attacked Iran. Helbig, who travelled to Iran as a graduate student, had even met Iranian President Mahmoud Ahmadinejad. Now back at the University of Virginia, in

Charlottesville, to complete her degree, Helbig describes a programme in disarray, in which social scientists — few of whom have regional or linguistic expertise — sat around for weeks at Fort Leavenworth in Kansas, with little in the way of region-specific training.

Matt Tompkins, Helbig’s fiancé and another human-terrain participant, describes other problems. As a PhD student in political science with a military background, he was assigned as a team leader in Baghdad; but the social scientist on his team had no relevant field-research experience, he says, and their de facto translator was a Moroccan who barely spoke English. As for the military commander they were supposed to be supporting, Tompkins says, “I didn’t get the inclination that he was particularly interested in what we were doing.”

McFate disputes the recruitment problems, although she says some academics have told her they fear being blackballed professionally if they work for the programme. Other supporters note that experiences of different teams have varied widely. Adam Silverman, a political scientist who works on a HTT outside of Baghdad, says he believes such work is valuable. “The programme is new, so it isn’t perfect,” he says. “It has growing pains.”

Working from what he describes as a mix of “unstructured interviews, casual dis-

cussions with members of the population, academic sources and the Internet”, Silverman has provided advice on everything from local funeral rites to agriculture. Although he is now working on oral histories, he acknowledges that his field research has been difficult to conduct. “We don’t interview anyone per se — we do try to talk with anyone who will talk with us,” he says. “I’ve had conversations with fish farmers, brickmakers, government officials and tribal leaders.”

However, it is not clear whether academic social scientists are even the key feature in successful human-terrain teams. McFate’s story about a team defusing the situation in Banat al Hassan was confirmed by Major Philip Carlson, who led the team in question. But the recommendation to let the man go wasn’t from a social scientist; it came from Carlson and an Iraqi-American analyst. There wasn’t even a social scientist on that team at the time.

McFate says that “smart, competent, well-trained people on a team” can be successful, as in this case, but that social scientists are needed to achieve the programme’s broader goals. But few, if any, definitive numbers exist by which to measure the programme’s effectiveness. Earlier this year, Colonel Martin Schweitzer, a military officer working in Afghanistan, testified before Congress that HTTs helped to reduce the number of operations involving military force in his region by 60–70%. Sceptical of those numbers, David Price, an anthropologist at Saint Martin’s University in Lacey, Washington, filed a Freedom of Information Act (FOIA) request to look at the report. Price says that what he got back was merely a correspondence stating the numbers; there was no

actual report. “When I got my FOIA reply I learned that there was no study out there substantiating any of this,” he says.

Even with the doubts surrounding the Human Terrain System, the Pentagon made another foray into the social sciences this April when Defense Secretary Robert Gates announced a broader military initiative. Called Project Minerva, it would fund work at universities that do research ranging from looking at Chinese military technology to Islamic radicalism.

Anthropologists critical of the Human Terrain System didn’t welcome Minerva either. In a 28 May letter to the White House’s Office of Management and Budget, the president of the AAA outlined a number of concerns,



**“Anthropology will thrive more as a discipline if the funding is not directly from the national security state.”**  
— Hugh Gusterson



## Lessons from the past

In the 1960s, as the United States faced an array of potential regional conflicts from southeast Asia to Latin America, its army began what seemed to be a modest project to examine the roots and causes of insurgency. Project Camelot, as it was called, would look at “the feasibility of developing a general social-systems model that would make it possible to predict and influence politically significant aspects of social change in the developing nations of the world”.

The seemingly innocuous project sparked a firestorm of criticism after researchers associated with

Camelot visited Chile, triggering media stories that the work was a prelude to US military involvement in the region. Camelot eventually became the most well known of the Vietnam-era social-science programmes, and has now become synonymous with much of the work from that time.

The Pentagon launched a number of similar social-science projects, setting off a pointed debate among anthropologists, who criticized the military for attempting to subvert social research for its own means and manipulate foreign cultures. Seymour Deitchman, a Pentagon

official who spearheaded many of the efforts, describes the rise, fall and backlash against these military social-science programmes in his 1976 book, *The Best-Laid Schemes: A Tale of Social Research and Bureaucracy*.

In one typical case in the late 1960s, a research group contacted an anthropologist about work it wanted to do for the Pentagon in the Congo. “The anthropologist immediately raised a storm,” says Deitchman, “writing to the American Anthropological Association and the press that an attempt was being made to enlist him in intelligence activities for the

suppression of Congo tribes in the conflict that was then in its final stages there.” In fact, the Pentagon had never agreed to fund the work.

Deitchman, who is now retired, sees many of the same frictions echoed in today’s efforts by the military to enlist social scientists. Although he does note that Congress and the secretary of defence support the modern studies — unlike in the Vietnam era — the underlying dynamics haven’t changed. “The ticking time bomb in government support of social science research is there,” he says, “just under the surface, waiting for the trigger.” **S.W.**

including the notion that having the Pentagon run such research creates a “potential conflict of interest”. Partly in response, the Pentagon forged a relationship with the National Science Foundation (NSF), which culminated earlier this year in the signing of a formal agreement. That, however, created new confusion, as many presumed that the foundation was cooperating on Minerva. Mark Weiss, director of the NSF’s behavioural and cognitive sciences division, insists that is not the case. “It is a Memorandum of Understanding that would allow for a number of different interactions that . . . would help enhance the flow of information from the social and behavioural sciences to the Department of Defense,” he says.

### Shopping for knowledge

One question concerns who would oversee the peer-review process for selecting grantees: the defence department or the NSF. Thomas Mahnken, the deputy assistant secretary of defense for policy planning, says that Minerva is budgeted for approximately \$100 million over five years, and that half that money would go through the NSF. The other half would go through the Pentagon, which he insists also has a well-tested peer-review process. “The two paths are complementary,” says Mahnken. “NSF certainly gives us access to a different pool of scholars.”

Critics of the programme, particularly anthropologists, point to a number of pitfalls associated with Minerva. One social scientist who works with the military warns of ‘ScamTechs’ — firms that are adept at getting defence department funding, regardless of the subject. And Hugh Gusterson, an anthropologist at George Mason University in Fairfax, Virginia, notes that a government contractor



**Soldiers of social science? US military meet tribal leaders in Ramadi to discuss Iraq’s reconstruction.**

recently contacted several colleagues, “shopping” for an anthropologist so that they could bid on Minerva, which requires university participation.

Another concern is that the Pentagon’s largesse could ultimately shift anthropologists away from their traditional role as advocates for the people and cultures they study. “Anthropologists ought to be involved [in the national security debate], but my fear is what makes anthropology appealing will be undercut and deformed if anthropologists are directly

answerable to the Pentagon in that conversation,” Gusterson says. “Anthropology will thrive more as a discipline if the funding is not directly from the national security state.”

Both the Pentagon and the NSF downplay any concerns that the defence department could flood the field with military funding. Weiss notes that NSF’s total annual budget for the behavioural and cognitive division is already about \$220 million. The Pentagon money, he says, is “not going to put us into a stratospheric level of funding”.

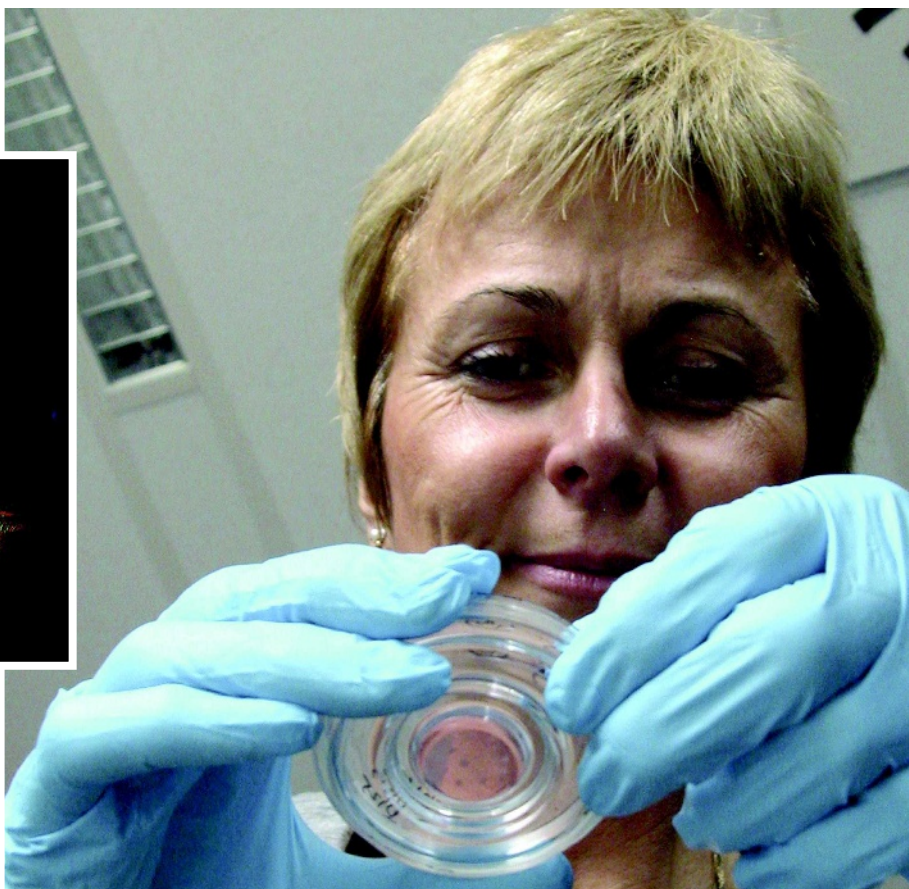
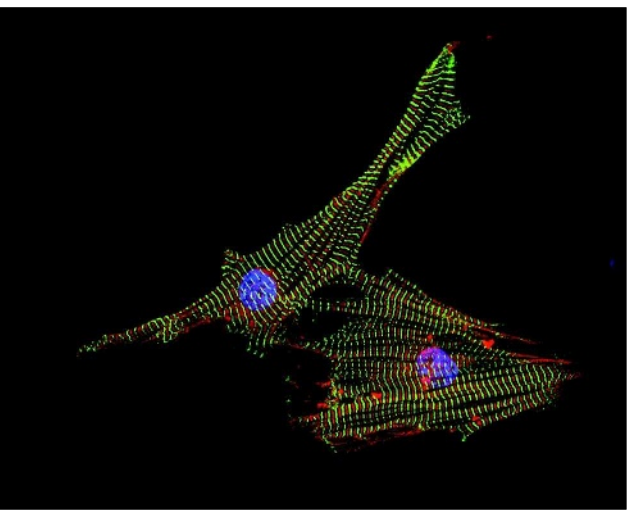
Meanwhile, researchers in other countries are grappling with some of the same issues. Two years ago, Britain’s Economic and Social Research Council was criticized for circumventing normal open academic competition by funding counterterrorism studies. Jeremy Keenan, a UK-based anthropologist and North Africa expert, says that the UK Foreign Office gave itself a respectable academic veneer by rerouting money quietly through the council. By contrast, “if one looks at the US military programme, it’s been very overt,” he says.

Other militaries have not yet developed an exact equivalent to the Human Terrain System, but they do have, on a smaller scale, social scientists providing advice to armed forces — they work in psychological operations units and provide training and education. And McFate says that some NATO allies have also expressed interest in setting up human-terrain-like programmes.

Whether other countries will be engulfed in the same controversy remains to be seen. McFate, for her part, puts the criticism down to a small but vocal group. “It’s just a very small segment of the anthropology community,” she says of the critics. “We’re not going to draft them.” ■

**Sharon Weinberger is a freelance writer.**

J. MOORE/GETTY IMAGES



A. BEQQALI

The first time Christine Mummery encountered a human heart, she didn't like what she saw. The vessels were brittle with atherosclerosis and the tissue was mottled like a bruise. But the biggest shock was that this diseased heart looked so different from that of a healthy mouse. How could she expect the stem-cell therapy she'd been testing on rodents to work in this alien-looking organ?

This is the type of thinking that led Mummery, a stem-cell scientist, to move to a hospital in May this year. In the six years since she saw the heart, she had become convinced that closer ties to clinicians and better access to human samples would make her research more applicable to patients. So she packed up her lab at the Hubrecht Institute in Utrecht, the Netherlands, and relocated to Leiden University Medical Center. "I thought we'd be better off in a clinical environment," she says. "It's so much easier if you can speak to clinicians over lunch."

Viviane Tabar also talks medicine at the table. A practicing neurosurgeon at Memorial Sloan-Kettering Cancer Center in New York, she spends two to three days a week caring for patients with brain tumours. But she thinks that surgery and the other tools that she has to help them aren't enough, so she is investigating the cells that might make brain therapies of the future. "The patient's perspective is often different from what a scientist might think," Tabar says. "They want to know what a technology can do for them on a very practical basis. So you learn to think much more pragmatically and to ask: would this really be helpful?"

Both Mummery and Tabar have veered from conventional career paths in hopes of making stem cells more useful therapeutically. They work in different countries and on different diseases. Mummery is well-established and Tabar is still early in her career. Yet both believe that

## BEING PATIENT

Cell therapies are as much about the patients as they are about the cells. **Monya Baker** meets two stem-cell scientists who have decided to put people first.

the best way to ensure that stem-cell therapies are 'translated' into patients is to be as close as possible to the patients themselves. "Some scientists could really benefit from just one or two days in the clinic," says Tabar.

Patients haven't always been the strongest focus for stem-cell researchers. Getting enough of the right cells to transplant was seen as the biggest stumbling block. That is changing with the derivation of human embryonic stem-cell lines, along with new techniques to make induced pluripotent stem (iPS) cells from patient skin biopsies. Researchers can now generate potentially limitless supplies of cells and use improved methods to grow them into the specialized types they want. The possibilities of human trials are edging nearer, and another problem is moving into focus: healing potential depends not just on the cells going into a patient, but on what they'll encounter inside.

Transplanted cells will be going into diseased bodies, says Marie Csete, scientific director of the California Institute for Regenerative Medicine (CIRM) in San Francisco, and it is the body as much as the cells that needs to

be studied if researchers hope to "predict cell behaviour when we put them into a distinctly pathological environment". But few people are well versed with both stem cells and patients. "We've had a crisis in embodying someone who really understands clinical medicine and who really understands basic science in one person," says Csete. There is no question that the field needs more people like Mummery and Tabar to chart a course between an interesting concept and ready-to-test therapy, says Fred Gage, who studies neural stem cells at the Salk Institute for Biological Studies in San Diego, California.

Mummery conveys a sense of purpose. She is tall, sometimes wears her short hair spiked up, and her speech and movements are brisk. She trained as a physicist at the University of Nottingham, UK, and when Mummery later moved into biology she thought that its practitioners should "measure something rather than just look".

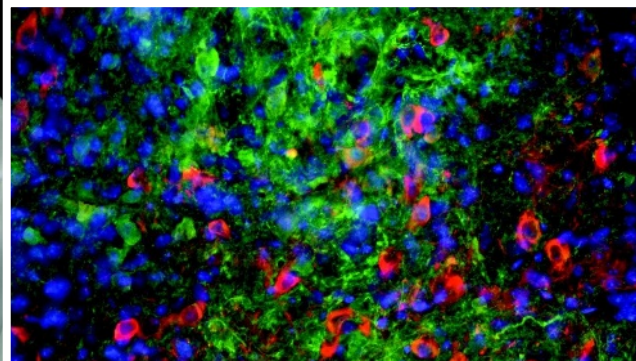
Mummery's quantitative approach has shown that, in some cases, stem-cell therapy may not live up to its initial promise even in animals. As a postdoc at the Hubrecht Institute she developed

J. LEWIS





Christine Mummery (opposite) works with cells for heart repair; Vivian Tabar (above left, with research technician Jayanthi Menon) studies cell therapy in the brain.



also took a postdoctoral fellowship in the laboratory of her husband, stem-cell biologist Lorenz Studer. The two had met collaborating on experiments in McKay's lab, and when Tabar went to complete her training, Studer established a lab at Sloan-Kettering studying how embryonic stem cells can create different sorts of neural cells in culture. Tabar wanted to develop animal models to test whether neurons differentiated from embryonic stem cells could restore function, and she wanted her own lab to do so. She established this in 2005.

Researchers sometimes worry that clinicians will rush to act on preliminary results: "Some clinicians want to try an idea tomorrow, rather than the day after," Mummery says. But Tabar says the opposite can be true: being a physician sometimes makes even the best research publications seem less exciting because they will rarely make a concrete difference in a patient's life. McKay recalls visiting Tabar's lab in New York, and her questioning the treatment implications of a recent paper on glioblastoma.

"Within five minutes, we were in another room, looking at patient records, talking about specific cases." It is precisely this "ability to dance on both sides of the aisle" that makes people like Tabar

able to ferret out the valuable experimental approaches, he says.

Some of Tabar and Studer's most recent work has addressed a pressing question: how closely cell transplants into the brain must be immunologically matched to the recipient. It has not been clear how much tissue rejection and inflammation could present a problem, because brain tissue is better protected from immune attack than that of other organs.

In a paper published this year, Tabar and Studer tackled this question in a mouse model of Parkinson's disease<sup>4</sup>. Drawing on the expertise of colleagues at the RIKEN Center for Developmental Biology in Kobe, Japan, they used therapeutic cloning to create embryonic stem-cell lines from 24 mice, differentiated

culturing techniques that induce embryonic stem cells to generate plentiful cardiomyocytes, the muscle cells that power the heart. But her later work has highlighted how difficult it could be to use these cells therapeutically.

A common way to track cells transplanted into a mouse, for example, is to engineer them to express a green fluorescent protein and then look for the tell-tale signal. But Mummery discovered that scar tissue within the heart also emits green fluorescence under certain conditions, and posited that at least some transplantation studies were mistaking scar tissue for successful engraftment<sup>1</sup>. She has also shown that cell-therapy results can be frustratingly short-lived. Other researchers showed that one month after derived cardiomyocytes are transfused into injured mouse hearts, the hearts pump significantly more blood<sup>2</sup>. Mummery showed that the effects disappear after three months<sup>3</sup>. "What her work has uncovered is that we need to pay attention to graft survival, appropriate alignment, and integration," says Ken Chien, a cardiologist and stem-cell biologist at Massachusetts General Hospital, Boston, who has collaborated with Mummery.

Around the time she glimpsed her first human heart in 2002, Mummery began to question the relevance of her work to humans. She realized that the way her lab was simulating heart attacks in healthy mice was a poor mimic for the real thing. As soon as a heart attack is diagnosed, heart surgeons work quickly to place stents in blocked vessels, opening them up again to allow oxygenated blood to perfuse the tissue. In mice, Mummery's group tied off blood vessels permanently (partly to comply with animal husbandry laws), making for

simpler, cleaner wounds. She recalls her discouragement: "You're already a year and hundreds of mice down the line, and the cardiologist would say 'that's not really it'."

Mummery was also becoming aware that clinicians are less concerned about the heart attack itself, and more about the ensuing heart failure. "I felt we needed to know more about the most relevant clinical problems," she says. She received recruitment offers from various institutions, including Harvard University. But in August 2007, she was offered the position at the hospital in Leiden — and she took it.

### Converging paths

Whereas Mummery has gone from basic research to medicine, Tabar has been diverted from medicine to the lab. In 1995, as a third-year neurosurgery resident at Memorial Sloan-Kettering, she found herself intrigued by a talk given by Ron McKay from the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, describing evidence for neural stem cells. "I had been taught that the brain is a post-mitotic organ; that everything dies and nothing regenerates," she says. "And his work was saying 'maybe not.'" Her clinical programme required a year of research, and she went to McKay's lab to study how to coax embryonic stem cells into forming dopaminergic neurons, the kind that die in Parkinson's disease.

Tabar finished her medical training and in 2002 she joined the medical faculty at Memorial Sloan-Kettering, where she specializes in brain cancers. But while seeing patients, she

**"It's so much easier if you can speak to clinicians over lunch."**  
— Christine Mummery

V. TABAR ET AL. NATURE MED. 14, 379–381 (2008).



these into dopamine-producing neurons, and transplanted them back into the animals. Mice that received cell transplants derived from their own cells improved; mice that received cells derived from other mice did not. The foreign transplants seemed to trigger an immune response allowing only a few cells to survive, suggesting that human cell transplants would need to be closely matched immunologically and perhaps even derived from the patients themselves.

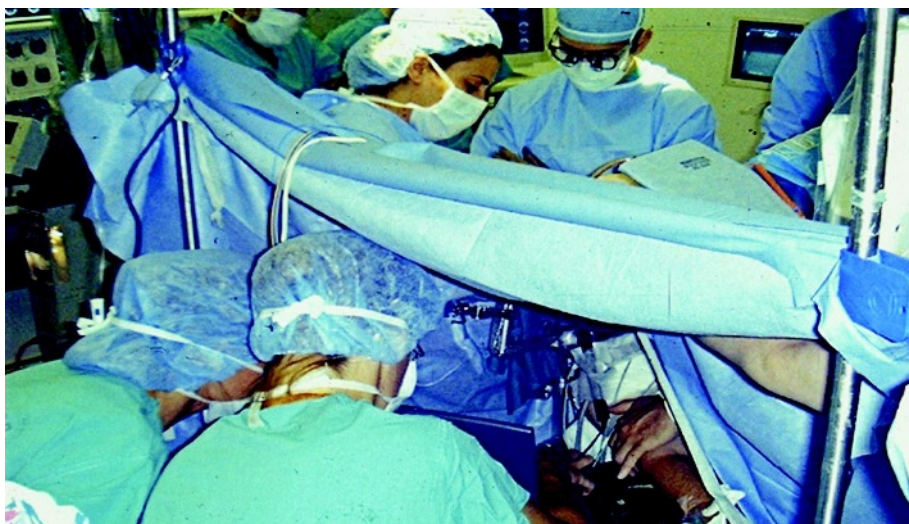
The results highlighted the difficulties with such therapies, but it was the paper's success that made headlines: it was the first time that cells made by therapeutic cloning were used to treat exactly the same animal from which they were derived. Tabar says it is difficult to gain recognition for the "unglamorous" work necessary to move from proof-of-principle research to a clinical application. "You could perhaps come up with a paradigm that works beautifully in animals in the lab and that puts together all these biological concepts," she says. "You can get it published, but bringing it to the patient will require a lot of mining through the details."

### The hard way

Academia does not tend to reward this type of detailed investigation; recognition is based on experimental 'firsts' and high-profile publications. If a scientist-clinician is not generating a stream of prestigious papers, then pressure increases to see patients, the activity that, after all, generates revenue for a physician's institution. Does Tabar ever consider how much simpler her life would be if she switched to all clinical or all scientific work? "Every day," she responds, laughing. But, she adds, she cannot imagine giving up either pursuit.

Tabar is now studying how cell replacement might help patients whose brains have been damaged by radiation therapy, as they would be after treatment for a brain tumour. Her lab administers various radiation regimes to rats, and then supplies cells — derived from embryonic stem cells — at various stages of differentiation and at various time points, trying to find the best combinations. She insists on measuring any benefits of these treatments using behavioural tests, rather than tissue integrity alone. "It is the clinical problem that I want to address rather than the simple histological or radiographic repair of the brain."

Stem-cell researchers are not alone in finding clinical research challenging: scientists



**Viviane Tabar (facing right, centre of picture) says treating patients helps her to ask research questions that are relevant to therapy.**

in almost every biomedical discipline are struggling to translate basic results into ones that can benefit patients. Stem-cell researchers are under particular pressure though. They must justify the massive investment made in them by funding agencies, philanthropies, and patients, such as the US\$3 billion of California taxpayer's money distributed by the CIRM.

Getting cell therapies to patients is also daunting because there is no established path to clinical approval. No treatments based on pluripotent stem cells have been approved for testing in humans, and earlier this year the US Food and Drug Administration halted plans by Californian biotech company Geron for the first trial in human patients of cells derived from embryonic stem cells. The risks for such therapies are almost impossible to assess, but the worry is that even a differentiated cell product could be unpredictable when administered to a patient, and might proliferate or transform into unwanted tissues.

Csete says that the best approach to translational research is funding collaborations or facilitating other practices that bring disease experts, clinicians and cell researchers together. The CIRM plans to announce several large, multi-year grants for such 'disease teams' later this year and has already funded just under two dozen 'planning grants' at around \$50,000 apiece to help collaborators at different institutions hammer out proposals.

The collaborative strategy seems to be working for Mummery, whose lab building is adjacent to the hospital at Leiden. There, she has access to fetal and adult human tissue and this has enabled her to try a new line of research: comparing human and mouse cardiac

cells. She had good collaborations in hospitals before, she says, but people seem more willing to help colleagues within a hospital than at an unrelated institution. Mummery also benefits from the clinicians' established procedures for gathering informed consent for tissue collection.

Mummery is using the human samples to work out which molecular markers are expressed when and where in developing hearts, and she has already found that rodents can sometimes lead scientists astray. In the mouse embryo, cardiac cells express distinct markers depending on whether they will form the atrial chambers that receive blood or the ventricular chambers that pump it out. But when Mummery's team probed human fetal tissue, they found that human ventricular cells — the type that weaken in heart failure and that cell therapy might replace — actually express the markers used to identify mouse atrial cells. "That made me realize that we are trying to make heart cells from human embryonic stem cells without really knowing what the cells are."

Even as she has moved closer to the clinicians who could do cell transplants, Mummery now feels that the timelines for those transplants are lengthening. She no longer thinks that pursuing cell transplants is the most productive use of her time. Instead, she is using them to screen for drugs that can change how the cells beat *in vitro*, in order to understand the cause and control of cardiac arrhythmias.

The patients are just next door, but Mummery is a bit less intent on putting cells into them. For now, she's decided, there is plenty that the cells can teach her in a dish. ■

**Monya Baker is the editor of Nature Reports Stem Cells.**



**"We've had a crisis finding someone who understands clinical medicine and basic science."**

— Marie Csete

1. van Laake, L. W. *et al.* *Stem Cell Research* **1**, 9–24 (2007).
2. Laffamme, M. A. *et al.* *Nature Biotechnol.* **25**, 1015–1024 (2007).
3. van Laake, L. W. *et al.* *Circ. Research* **102**, 1008–1010 (2008).
4. Tabar, V. *et al.* *Nature Med.* **14**, 379–381 (2008).

## CORRESPONDENCE

## Don't release other people's data without their consent

SIR — I am astounded by the audacity of someone photographing the presentation of another researcher and then publishing their data without the presenter's permission ('Physicists aflutter about data photographed at conference' *Nature* **455**, 7; 2008). In what scientific forum, other than apparently the arXiv.org preprint server, is that permissible practice?

The implication in your News story that videorecording conference proceedings somehow justifies this unfortunate incident is misleading. The point is not that data are recorded through a particular medium. We all take notes at meetings in some form, whether physical or mental, and who hasn't had their poster transcribed nearly verbatim? Rather, the issue is whether the information is released in a fair and representative manner.

Although, as you note, videotaping conference proceedings is common in biology, we operate under an implicit, and often explicit, ethic that data presented at meetings are personal communications. As such, publication of personal communications by a second party requires formal approval from the originating researchers. This practice strikes a balance between the public good that arises from collegial sharing of preliminary results and preservation of investigators' rights to ownership of their intellectual work. Therefore, except under exceptional circumstances, scientists ought to obtain permission to cite the unpublished works of others. Sometimes investigators may unfairly withhold data that are so critical that they justify overlooking what is the norm in most academic communities. But any such putatively exceptional case of data release should



require unequivocal justification.

Ethics aside, what exactly was the purpose of reporting incompletely vetted, possibly erroneous experimental results? Can the group who released the data provide assurance that the information gleaned during the presentation adequately represents the original data in all its potential complexity? If not, there seems to be little justification for, or value in, usurping the intellectual rights of the group that originally generated and presented the results. And it is doubtful whether all those who contributed to the project received proper credit.

This case violates the spirit of collegiality that most scientists hold as an ideal in our public discourse. We all accept that others may scoop our work. We should not have to worry about being scooped by our own data.  
**Daniel N. Frank** Molecular, Cellular and Developmental Biology, Mucosal and Vaccine Research Program Colorado, University of Colorado–Boulder, Boulder, Colorado 80309, USA  
e-mail: daniel.frank@colorado.edu

## Further reflections on how we interpret the actions of others

SIR — In their Essay 'Behind the looking-glass' (*Nature* **454**, 167–168; 2008), Antonio Damasio and Kaspar Meyer

suggest how mirror neurons might work. But they need to reflect on other aspects of the mirror phenomenon to complete the picture.

Mirror neurons are known for their intriguing property of discharging when a particular motor act is either being performed or being observed. Damasio and Meyer describe them as neural ensembles in higher-order association areas called CDZs (for 'convergence–divergence zones') that collect information from specific sensory areas and signal back to those areas. Action understanding (as in the authors' example of hearing a peanut being cracked) depends on activation of this network.

The CDZ model attempts to explain the mechanism underlying action understanding. But it overlooks a fundamental feature of the mirror mechanism: that is, the capacity to transform sensory information into a motor format — why should we have a copy of the actions of others in our motor system?

We can certainly recognize biological actions using sensory information and performing the kind of processing suggested by the CDZ model. But mirror neurons indicate that we must also have another mechanism for understanding another's actions. That mechanism directly maps sensory information on cortical motor neurons, providing the observer with an immediate

representation of the motor acts being performed by others. There is no need for a higher-order association, as the CDZ model requires.

This, of course, does not imply that mirror neurons alone 'understand' the actions of others. Such an interpretation of the mirror system would go against all we know about the complexity of cortical organization. The point at issue is the specific contribution of mirror neurons to action understanding. Because of their motor nature, these neurons add a new, personal dimension to our capacity for understanding others that is based on our own motor knowledge and experience.

So, in spite of its heuristic value, the CDZ model underestimates the motor aspect of the mirror mechanism. It was this mechanism that prompted the description of action understanding as "the result of a 'first-person' process where the self feels like an actor, rather than a spectator" (M. Jeannerod *The Cognitive Neuroscience of Action*, Blackwell, 1997).

**Giacomo Rizzolatti** Dipartimento di Neuroscienze, Università di Parma, Via Volturno 39, 43100 Parma, Italy  
e-mail: giacomo.rizzolatti@unipr.it  
**Corrado Sinigaglia** Dipartimento di Filosofia, Università degli Studi di Milano, Via Festa del Perdono 7, 20122 Milano, Italy

## Austria: Academy of Sciences states its case

SIR — As president and secretary-general of the Austrian Academy of Sciences, we wish to clarify the academy's position in the investigation into the alleged scientific misconduct associated with the urological clinical trial that you discuss in your Editorial 'Scandalous behaviour' (*Nature* **454**, 917–918; 2008). Contrary to your implications, the academy is committed to help in the resolution of this case.

You note that the academy put

D. PARKINS



## CORRESPONDENCE

## Don't release other people's data without their consent

SIR — I am astounded by the audacity of someone photographing the presentation of another researcher and then publishing their data without the presenter's permission ('Physicists aflutter about data photographed at conference' *Nature* **455**, 7; 2008). In what scientific forum, other than apparently the arXiv.org preprint server, is that permissible practice?

The implication in your News story that videorecording conference proceedings somehow justifies this unfortunate incident is misleading. The point is not that data are recorded through a particular medium. We all take notes at meetings in some form, whether physical or mental, and who hasn't had their poster transcribed nearly verbatim? Rather, the issue is whether the information is released in a fair and representative manner.

Although, as you note, videotaping conference proceedings is common in biology, we operate under an implicit, and often explicit, ethic that data presented at meetings are personal communications. As such, publication of personal communications by a second party requires formal approval from the originating researchers. This practice strikes a balance between the public good that arises from collegial sharing of preliminary results and preservation of investigators' rights to ownership of their intellectual work. Therefore, except under exceptional circumstances, scientists ought to obtain permission to cite the unpublished works of others. Sometimes investigators may unfairly withhold data that are so critical that they justify overlooking what is the norm in most academic communities. But any such putatively exceptional case of data release should



require unequivocal justification.

Ethics aside, what exactly was the purpose of reporting incompletely vetted, possibly erroneous experimental results? Can the group who released the data provide assurance that the information gleaned during the presentation adequately represents the original data in all its potential complexity? If not, there seems to be little justification for, or value in, usurping the intellectual rights of the group that originally generated and presented the results. And it is doubtful whether all those who contributed to the project received proper credit.

This case violates the spirit of collegiality that most scientists hold as an ideal in our public discourse. We all accept that others may scoop our work. We should not have to worry about being scooped by our own data.  
**Daniel N. Frank** Molecular, Cellular and Developmental Biology, Mucosal and Vaccine Research Program Colorado, University of Colorado–Boulder, Boulder, Colorado 80309, USA  
e-mail: daniel.frank@colorado.edu

## Further reflections on how we interpret the actions of others

SIR — In their Essay 'Behind the looking-glass' (*Nature* **454**, 167–168; 2008), Antonio Damasio and Kaspar Meyer

suggest how mirror neurons might work. But they need to reflect on other aspects of the mirror phenomenon to complete the picture.

Mirror neurons are known for their intriguing property of discharging when a particular motor act is either being performed or being observed. Damasio and Meyer describe them as neural ensembles in higher-order association areas called CDZs (for 'convergence–divergence zones') that collect information from specific sensory areas and signal back to those areas. Action understanding (as in the authors' example of hearing a peanut being cracked) depends on activation of this network.

The CDZ model attempts to explain the mechanism underlying action understanding. But it overlooks a fundamental feature of the mirror mechanism: that is, the capacity to transform sensory information into a motor format — why should we have a copy of the actions of others in our motor system?

We can certainly recognize biological actions using sensory information and performing the kind of processing suggested by the CDZ model. But mirror neurons indicate that we must also have another mechanism for understanding another's actions. That mechanism directly maps sensory information on cortical motor neurons, providing the observer with an immediate

representation of the motor acts being performed by others. There is no need for a higher-order association, as the CDZ model requires.

This, of course, does not imply that mirror neurons alone 'understand' the actions of others. Such an interpretation of the mirror system would go against all we know about the complexity of cortical organization. The point at issue is the specific contribution of mirror neurons to action understanding. Because of their motor nature, these neurons add a new, personal dimension to our capacity for understanding others that is based on our own motor knowledge and experience.

So, in spite of its heuristic value, the CDZ model underestimates the motor aspect of the mirror mechanism. It was this mechanism that prompted the description of action understanding as "the result of a 'first-person' process where the self feels like an actor, rather than a spectator" (M. Jeannerod *The Cognitive Neuroscience of Action*, Blackwell, 1997).

**Giacomo Rizzolatti** Dipartimento di Neuroscienze, Università di Parma, Via Volturno 39, 43100 Parma, Italy  
e-mail: giacomo.rizzolatti@unipr.it  
**Corrado Sinigaglia** Dipartimento di Filosofia, Università degli Studi di Milano, Via Festa del Perdono 7, 20122 Milano, Italy

## Austria: Academy of Sciences states its case

SIR — As president and secretary-general of the Austrian Academy of Sciences, we wish to clarify the academy's position in the investigation into the alleged scientific misconduct associated with the urological clinical trial that you discuss in your Editorial 'Scandalous behaviour' (*Nature* **454**, 917–918; 2008). Contrary to your implications, the academy is committed to help in the resolution of this case.

You note that the academy put

D. PARKINS



**“‘Social butterflies’ and ‘open skies’ are equally important for the transmission of HIV.”** Karunesh Tuli, page 594

its investigation, requested by the rector of the Medical University of Innsbruck, on hold following the announcement by the university council that the rector would shortly lose his position (which has since happened). However, the investigations mandated by the rector are still under way, although no externally visible measures have yet been carried out. His withdrawal will not terminate or otherwise influence the investigations.

The academy has declined to make any pronouncement in advance of a verdict by a formal enquiry because the affair's scientific, ethical, legal and political issues must all be taken into consideration. A premature statement would not help to clarify the situation and would encourage accusations that its release was politically motivated.

We consider that it is of the utmost importance for the scientific community in Austria to investigate the case in an impartial and unprejudiced way, which will allow us to draw valid and independent conclusions. Individuals being chosen for the investigation will be completely independent and selected for their high scientific competence and moral integrity.

We therefore strongly reject your implication that the Austrian Academy of Sciences could be directly or indirectly involved in any political moves that might promote scientific misconduct and corrupt the scientific community.

**Peter Schuster, Herwig Friesinger**  
Austrian Academy of Sciences,  
Dr Ignaz Seipel-Platz 2, 1010 Vienna,  
Austria  
e-mail: peter.schuster@oeaw.ac.at

## Science journals have been slow to make themselves audible

SIR — Podcasting has become very popular, mainly as a medium for entertainment. But it also holds huge potential for the visually impaired and others, such

as dyslexics, who have reading difficulties. Simultaneously reading and listening to read-aloud news articles and scientific papers, for example, could increase readers' concentration and absorption of information. Such audio files would open a new world for the blind or partially sighted.

Software is available that translates text from digital files or directly from the Internet into a listener-friendly audio file, but it is expensive. Some freeware has built-in 'read out-loud' functions, but the quality is generally inferior.

Several newspapers and magazines already offer subscribers podcasts containing complete and navigable issues in read-aloud format. But the scientific press seems to be lagging behind. The *Nature* podcasts are a good start, but when shall we be able to listen to sections such as Research Highlights, News and Correspondence as downloadable audio?

**Wouter M. J. Achten** Division Forest,  
Nature and Landscape, Katholieke  
Universiteit Leuven, Celestijnenlaan  
200E Box 2411, 3001 Leuven, Belgium  
e-mail: wouter.achten@ees.kuleuven.be

## Don't forget people and specimens that make the database

SIR — Further to points raised in your Feature 'The future of biocuration' (*Nature* **455**, 47–50; 2008), an example of the inadequate state of biocuration is to be found in the large number of entries in GenBank listed as "unpublished". In many cases, a quick online search of journal listings turns up the publication. Obviously, the journals and GenBank are not communicating as well as they ought.

It's also important not to lose sight of the underlying need to curate biological specimens and materials, a function that needs much more support. Biology deals with actual organisms, so proper curation of voucher specimens

and reference cultures, or their equivalent, is essential to confirm, test and build on previous studies.

There is also a lack of support for many of those taking time to build up data sets. "I spent lots of time online editing a database" doesn't get you anywhere on a resumé or tenure review, or help an unpaid volunteer make a living.

**David Campbell** 425 Scientific  
Collections Building, Department of  
Biological Sciences, Biodiversity and  
Systematics, University of Alabama,  
Box 870345, Tuscaloosa, Alabama  
35487-0345, USA  
e-mail: amblema@bama.ua.edu

## Religion and science: a guide for the 'perplexed'

SIR — There is no need for Matthew Cobb and Jerry Coyne (*Nature* **454**, 1049; 2008) to be "perplexed" by your Editorial concerning the funding of science and religion ('Templeton's legacy' *Nature* **454**, 253–254; 2008). As their Correspondence implies, the scientific study of religion is itself an important topic, and the Templeton Foundation gives grants for such work, for example in the field of cognitive psychology and the evolution of religious belief.

There are many reasons why the funding of academic research in this arena should be supported. Far from being in "fundamental conflict", history shows that there has been a constant traffic of ideas between science and religion, which provide complementary accounts of the same reality. In Stephen Hawking's colourful words, religion addresses the question "Why does the Universe go to all the bother of existing?". Boundary disputes arise when science claims too much (as in the philosophy of 'scientism') or when religion encroaches on science (as in so-called intelligent design, or creationism).

One pragmatic reason for supporting good academic

science–religion research is that most of the world's taxpayers, who fund science, have religious beliefs. Pitting science against religion in that context is not a smart move for the future of science.

Next year is the bicentenary of Darwin's birth and also marks 150 years since the publication of his *On the Origin of Species*. It would be great if atheists, agnostics and religious believers alike could celebrate Darwin as the brilliant biologist he was, not as the icon of a particular ideology.

**Denis R. Alexander** The Faraday  
Institute, St Edmund's College,  
Cambridge CB3 0BN, UK  
e-mail: dra24@hermes.cam.ac.uk

## Religion and science: separated by an unbridgeable chasm

SIR — In his Correspondence 'Religion: science is partially based on faith' (*Nature* **455**, 26–27; 2008), Jonathan Cowie argues that science and religion are more similar than often thought, suggesting that experimental application of the scientific method involves faith. However, he is conflating two different meanings of 'faith'.

Cowie's inherent definition of faith pertains to scientists' hopeful expectations that experiments will verify their (rational) hypotheses, whereas the definition relevant to religion is belief without evidence.

Insisting on evidence-based beliefs separates science starkly from religion. Contrary to Cowie's assertion and to the goals of the Templeton Foundation, the chasm between science and religion is fundamentally unbridgeable.

**Peter Wigley** School of Pharmacy  
and Medical Sciences, Reid Building,  
University of South Australia, Adelaide,  
South Australia 5000, Australia  
e-mail: peter.wigley@unisa.edu.au

Readers are welcome to comment at  
<http://tinyurl.com/5rgnuc>

**“‘Social butterflies’ and ‘open skies’ are equally important for the transmission of HIV.” Karunesh Tuli, page 594**

its investigation, requested by the rector of the Medical University of Innsbruck, on hold following the announcement by the university council that the rector would shortly lose his position (which has since happened). However, the investigations mandated by the rector are still under way, although no externally visible measures have yet been carried out. His withdrawal will not terminate or otherwise influence the investigations.

The academy has declined to make any pronouncement in advance of a verdict by a formal enquiry because the affair's scientific, ethical, legal and political issues must all be taken into consideration. A premature statement would not help to clarify the situation and would encourage accusations that its release was politically motivated.

We consider that it is of the utmost importance for the scientific community in Austria to investigate the case in an impartial and unprejudiced way, which will allow us to draw valid and independent conclusions. Individuals being chosen for the investigation will be completely independent and selected for their high scientific competence and moral integrity.

We therefore strongly reject your implication that the Austrian Academy of Sciences could be directly or indirectly involved in any political moves that might promote scientific misconduct and corrupt the scientific community.

**Peter Schuster, Herwig Friesinger**  
Austrian Academy of Sciences,  
Dr Ignaz Seipel-Platz 2, 1010 Vienna,  
Austria  
e-mail: peter.schuster@oeaw.ac.at

## Science journals have been slow to make themselves audible

SIR — Podcasting has become very popular, mainly as a medium for entertainment. But it also holds huge potential for the visually impaired and others, such

as dyslexics, who have reading difficulties. Simultaneously reading and listening to read-aloud news articles and scientific papers, for example, could increase readers' concentration and absorption of information. Such audio files would open a new world for the blind or partially sighted.

Software is available that translates text from digital files or directly from the Internet into a listener-friendly audio file, but it is expensive. Some freeware has built-in 'read out-loud' functions, but the quality is generally inferior.

Several newspapers and magazines already offer subscribers podcasts containing complete and navigable issues in read-aloud format. But the scientific press seems to be lagging behind. The *Nature* podcasts are a good start, but when shall we be able to listen to sections such as Research Highlights, News and Correspondence as downloadable audio?

**Wouter M. J. Achten** Division Forest,  
Nature and Landscape, Katholieke  
Universiteit Leuven, Celestijnenlaan  
200E Box 2411, 3001 Leuven, Belgium  
e-mail: wouter.achten@ees.kuleuven.be

## Don't forget people and specimens that make the database

SIR — Further to points raised in your Feature 'The future of biocuration' (*Nature* **455**, 47–50; 2008), an example of the inadequate state of biocuration is to be found in the large number of entries in GenBank listed as "unpublished". In many cases, a quick online search of journal listings turns up the publication. Obviously, the journals and GenBank are not communicating as well as they ought.

It's also important not to lose sight of the underlying need to curate biological specimens and materials, a function that needs much more support. Biology deals with actual organisms, so proper curation of voucher specimens

and reference cultures, or their equivalent, is essential to confirm, test and build on previous studies.

There is also a lack of support for many of those taking time to build up data sets. "I spent lots of time online editing a database" doesn't get you anywhere on a resumé or tenure review, or help an unpaid volunteer make a living.

**David Campbell** 425 Scientific  
Collections Building, Department of  
Biological Sciences, Biodiversity and  
Systematics, University of Alabama,  
Box 870345, Tuscaloosa, Alabama  
35487-0345, USA  
e-mail: amblema@bama.ua.edu

## Religion and science: a guide for the 'perplexed'

SIR — There is no need for Matthew Cobb and Jerry Coyne (*Nature* **454**, 1049; 2008) to be "perplexed" by your Editorial concerning the funding of science and religion ('Templeton's legacy' *Nature* **454**, 253–254; 2008). As their Correspondence implies, the scientific study of religion is itself an important topic, and the Templeton Foundation gives grants for such work, for example in the field of cognitive psychology and the evolution of religious belief.

There are many reasons why the funding of academic research in this arena should be supported. Far from being in "fundamental conflict", history shows that there has been a constant traffic of ideas between science and religion, which provide complementary accounts of the same reality. In Stephen Hawking's colourful words, religion addresses the question "Why does the Universe go to all the bother of existing?". Boundary disputes arise when science claims too much (as in the philosophy of 'scientism') or when religion encroaches on science (as in so-called intelligent design, or creationism).

One pragmatic reason for supporting good academic

science–religion research is that most of the world's taxpayers, who fund science, have religious beliefs. Pitting science against religion in that context is not a smart move for the future of science.

Next year is the bicentenary of Darwin's birth and also marks 150 years since the publication of his *On the Origin of Species*. It would be great if atheists, agnostics and religious believers alike could celebrate Darwin as the brilliant biologist he was, not as the icon of a particular ideology.

**Denis R. Alexander** The Faraday  
Institute, St Edmund's College,  
Cambridge CB3 0BN, UK  
e-mail: dra24@hermes.cam.ac.uk

## Religion and science: separated by an unbridgeable chasm

SIR — In his Correspondence 'Religion: science is partially based on faith' (*Nature* **455**, 26–27; 2008), Jonathan Cowie argues that science and religion are more similar than often thought, suggesting that experimental application of the scientific method involves faith. However, he is conflating two different meanings of 'faith'.

Cowie's inherent definition of faith pertains to scientists' hopeful expectations that experiments will verify their (rational) hypotheses, whereas the definition relevant to religion is belief without evidence.

Insisting on evidence-based beliefs separates science starkly from religion. Contrary to Cowie's assertion and to the goals of the Templeton Foundation, the chasm between science and religion is fundamentally unbridgeable.

**Peter Wigley** School of Pharmacy  
and Medical Sciences, Reid Building,  
University of South Australia, Adelaide,  
South Australia 5000, Australia  
e-mail: peter.wigley@unisa.edu.au

Readers are welcome to comment at  
<http://tinyurl.com/5rgnuc>

## COMMENTARY

## HIV immunology needs a new direction

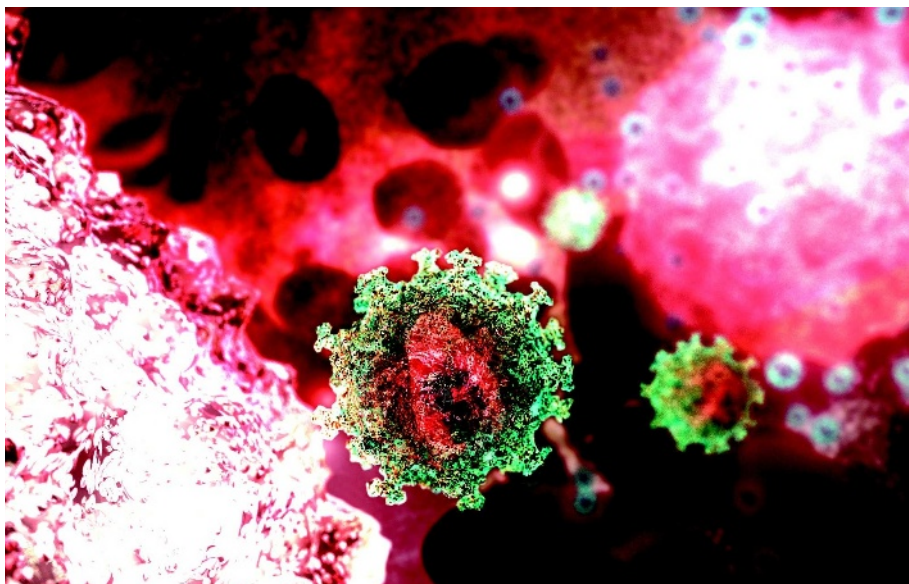
Researchers need to get past the standard model of vaccine development and focus on how immune responses are specifically tailored to retroviruses, argue **Ruslan Medzhitov** and **Dan Littman**.

**R**ecent advances in immunology and failures in HIV-vaccine development suggest that it is time to rethink the current approach to developing an HIV vaccine. Better communication and cooperation is needed between vaccinologists, virologists and the growing number of researchers studying innate immune responses. HIV has evaded the attempts of vaccinologists for 25 years partly because of the unique characteristics of the virus, including its extremely high mutation rate, which enables immune evasion, and its ability to infect and deplete the major orchestrators of the immune response — the CD4<sup>+</sup> T cells. But, in our opinion, there are other reasons for failure.

Although vaccination arguably represents biomedical science's greatest triumph over disease, so far the successes can largely be ascribed to empirical findings, scattershot approaches that have yielded great dividends for scourges such as smallpox and polio, but little to speak of for others including HIV and malaria. This reflects our current, clearly limited, level of understanding of the immune system. Specifically, the 'rules' for making a successful vaccine are currently unknown.

If there is any guiding principle to this formidable task, it is that a vaccine should mimic, as much as possible, the immune recognition events that happen during a natural infection with the same pathogen. These would be likely to engage the appropriate mechanisms of immune protection. This is why attenuated pathogens are the most successful vaccines currently available. For HIV, the use of attenuated strains is not an option at present. We therefore need to understand how the virus is normally sensed by the immune system.

Our understanding of early host responses, although incomplete, has improved dramatically in the past decade owing to a focus on the innate immune system, a first responder to pathogens that induces antimicrobial defence mechanisms and activates adaptive immunity. The human immune system has to deal with a variety of pathogens, ranging from RNA viruses to 30-foot-long tapeworms. Naturally, the ways in which these intruders are sensed and handled vary. Indeed several families of microbial sensors have now



Understanding how cells initially sense and recognize HIV could aid vaccine development (computer art).

been identified that detect different classes of pathogens and trigger activation of adaptive immunity by different mechanisms<sup>1,2</sup>.

Major gaps remain in the understanding of how HIV and retroviral infection in general is sensed by the innate immune system, and how this initial sensing is translated into the activation of adaptive immunity. In our view, there is currently insufficient effort being devoted to addressing these gaps and this is irreconcilable with the urgent need for a vaccine.

Studying HIV is of course complicated by its human-specific tropism, which precludes many definitive experiments. Although the mouse, as a model system, is often dismissed as irrelevant to HIV infection, there are likely to be common basic principles of retroviral immune recognition that need to be defined. Innate sensors may recognize features of replication that are shared among retroviruses but no other classes of viruses. For example, newly reverse-transcribed retroviral DNA may activate a specialized signalling pathway. Further study of the numerous host factors co-opted by HIV for its replication<sup>3</sup> might provide insight, as these may recruit antiviral innate sensors that limit viral spread.

Approaches currently used in HIV vaccine

development are largely based on immunological paradigms wrought from studies using antigen immunizations and infections with several model pathogens. These paradigms may not apply to retroviral infections. In the absence of a basic understanding of how the host immune system responds to retroviral infection, the approaches tried so far have been more or less random, and unsuccessful, attempts to see what might work. Most of the resources devoted to HIV vaccine development have been spent on such efforts, which may remain futile until we gain some understanding of the basic mechanisms of retroviral recognition and induction of adaptive immune responses. ■

**Ruslan Medzhitov** is professor of immunobiology at Yale University School of Medicine, 300 Cedar Street, New Haven, Connecticut 06520, USA. e-mail: [ruslan.medzhitov@yale.edu](mailto:ruslan.medzhitov@yale.edu)

**Dan Littman** is professor of molecular immunology and professor of pathology and microbiology at New York University School of Medicine, 550 First Avenue, New York 10016, USA. e-mail: [dan.littman@med.nyu.edu](mailto:dan.littman@med.nyu.edu)

1. Pichlmair, A. & Reis e Sousa, C. *Immunity* **27**, 370–383 (2007).

2. Ishii, K.J., Koyama, S., Nakagawa, A., Coban, C. & Akira, S. *Cell Host Microbe* **3**, 352–363 (2008).

3. Brass, A.L. *et al. Science* **319**, 921–926 (2008).

Discuss this Commentary online at <http://tinyurl.com/4683h3>.

**"So far, approaches have been more or less random attempts to see what might work."**



## COMMENTARY

## UK physics gets a health check

The field is healthy, says **Bill Wakeham**, but scientists need to reclaim the intellectual ownership of research at the margins of the discipline such as medical or atmospheric physics.

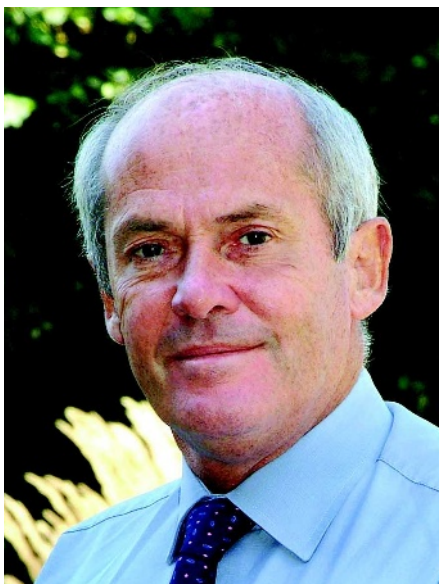
Physics research funding in the United Kingdom has been under the spotlight in the past year. In 2007, a restructuring of the research councils and a government budget settlement were perceived by the community as being unfavourable to the discipline. This prompted a wide review of the health of UK physics, looking at its international status, funding arrangements, university provision, school education, careers and skill supply. I was invited by the research councils to chair the review panel, drawing on expertise across the discipline and from overseas. Our report was published on 1 October.

UK physics is strong but faces important challenges. International reviews conducted in 2000 and 2005 recognized its outputs in terms of publications and citations. An analysis based on ISI citation data from 1998–2008 places the combination of UK physics and space science second only to that of the United States. The study we commissioned from Leiden University in the Netherlands showed that despite having a lower publication volume than Japan, France and Germany, the United Kingdom receives a higher citation rate per publication.

The discipline's impact on the wider economy is also strong. An Institute of Physics study calculated that the economic activity of physics-based sectors in the United Kingdom was £70 billion (US\$125 billion) in 2005. These include information technology, construction, transport and aerospace. Physics graduates also find employment in sectors with no obvious connection to the subject, notably finance.

The United Kingdom's academic physics workforce is stable, with around 4,000 academic and postdoctoral researchers. The annual undergraduate intake has increased from 3,415 in 2002/03 to 3,885 in 2006/07, so the appeal of the subject to a core group is clearly robust. Research in the United Kingdom has benefited from an 82% increase in the overall science and research budget in the five years to 2006/07, and physics-research funding has increased by 34% during this period, in line with the growth of UK gross domestic product.

Despite this superficial picture of stability, physics in the United Kingdom has seen significant changes during the past 20 years, with a number of departments closing



**Bill Wakeham:** undoing the merger of research councils would be rash.

or merging. Factors driving this change include the expansion of the university sector to embrace the former polytechnics, the increased focus of universities' research strategies driven by the Research Assessment Exercise and the reduction in the unit resource for undergraduate physics teaching.

Departmental focus on particular areas of physics has had a significant effect. Increasingly, physics departments rely on the core physics programmes of the Engineering and Physical Sciences Research Council and the Science and

Technology Facilities Council (STFC); the latter was formed last year through the merger of the Particle Physics and Astronomy Research Council and the Council for the

Central Laboratory of the Research Councils. Other British research councils invest in physics — notably the Medical Research Council, the Natural Environment Research Council and the Biotechnology and Biological Sciences Research Council — but 95% of their physics-related funding is spent in non-physics departments.

This tells us three things: 1) the dependence on a small number of funding streams makes physics departments vulnerable; 2) physics departments have not fully grasped the opportunities in interdisciplinary research and

some subdisciplines; and 3) the intellectual leadership of research at the margins of the discipline is often in other departments. There may be reasons for this at an individual institutional level, but at a national level it militates against the ability of the discipline to work with large sections of industry. Physics departments need to reclaim the intellectual ownership of some parts of their discipline and they need support from the funding agencies to achieve this.

Like many developed countries, the United Kingdom faces a challenge to maintain the supply and recruitment of research scientists. The problem is acute in physics and one that my panel explored at several points of the pipeline. Government intervention may be necessary to solve some of the problems. Physics needs to be taught by those trained in the subject. Too many young people are being taught physics by teachers with little relevant background. Worrying also is the decline in uptake of physics by girls, which has fallen by 16% compared with a 12% fall for boys in the four-year period assessed.

The merger of two research councils to form the STFC was generally welcomed by the research community. It would be precipitate to undo this process rapidly and we are not convinced that significant changes to the remits of individual research councils to unify physics funding or to separate facility and grant spend would be helpful. Nevertheless, the community must be confident that the current remits serve the best interests of UK research and physics and in particular, that commitments to facilities do not have an undesirable impact on the availability of grant funding. We think that government should revisit the current arrangement at a suitable time, guided by the more detailed review of STFC structures that has been commissioned.

The value of physics to the United Kingdom is such that weaknesses in its structures and in the skills pipeline must be addressed. Some aspects of this problem have been recognized, but action now needs to be pursued with greater urgency.

**Bill Wakeham** is chair of the Research Councils UK Review of Physics. He is a member of the Council of the Engineering and Physical Sciences Research Council and vice-chancellor of the University of Southampton, University Road, Southampton SO17 1BJ, UK.

e-mail: vice-chancellor@soton.ac.uk

Discuss this Commentary online at <http://tinyurl.com/4hx8zq>.

**"95% of physics-related funding is spent in non-physics departments."**

## BOOKS &amp; ARTS

# A fluid approach to HIV

A readable anthropological account of social networks in South Africa and Uganda explains differences in the spread of sexually transmitted diseases in those countries, finds **Karunesh Tuli**.

## Unimagined Community: Sex, Networks, and AIDS in Uganda and South Africa

by Robert J. Thornton

University of California Press: 2008.

282 pp. \$60 (hbk), \$24.95 (pbk)

"Abstain," said our hostess as we got off the bus to spend a week in Zwelethemba, a township in Western Cape, South Africa. She added that for her grown-up children, her message is different: "Be wise, condomize." A few days later, our group of students and teachers, visiting the country from the United States to learn about how South Africans handle health problems, was interviewed by some local journalists. After their questions dried up, we asked them what was the biggest problem facing their community. "AIDS," they replied.

Stories about AIDS and violence are often part of the same narrative in South Africa. In *Unimagined Community*, Robert Thornton describes the 2006 rape trial of Jacob Zuma, South Africa's deputy president until 2005. The Johannesburg High Court found him not guilty but the hearings provided Thornton with fertile material for anthropological analysis. Zuma's accuser, a family friend and an AIDS activist, was HIV-positive. Zuma told the court he had unprotected sex with her even though he knew her HIV status; he followed it with a shower to prevent infection.

*Unimagined Community* is the latest publication in the California Series in Public Anthropology. Robert Borofsky, the series editor, challenges anthropologists to write for a popular audience, inspired by Jared Diamond's much-read *Guns, Germs, and Steel* and Anne Fadiman's *The Spirit Catches You and You Fall Down*. Both writers used anthropological approaches but lack formal training in the field. Thornton reaches out to a general readership with a fresh interpretation of a pressing social problem. With his vigorous and imaginative writing, he succeeds admirably.

Thornton suggests that, in South Africa, sex is more than a personal quest for pleasure. It can be viewed community-wide as an exchange of fluids between males and females, part of a broader flow of objects and services between sexual partners. The men and women are hubs in large sexual networks that are regional and even countrywide. HIV follows the same paths as the flow of sexual fluids.

Traditional healers are reluctant to prescribe



Despite safe-sex campaigns, AIDS is harder to prevent in South African communities than in Uganda.

condoms. Thornton attended a workshop on HIV/AIDS for healers where the facilitator asked them to recommend condoms to their clients. Yet the attendees assembled a formidable case against the advice, saying, for example, that condoms could lead to illness in the man by causing a backup of semen, or could slip off and disappear inside his partner. Many South Africans believe that damming the flow of fluids disrupts a vital exchange and damages social networks.

Repelling a woman's sexual advances is akin to denying her "human right to sex", as Zuma declared in his defence, claiming his accuser had enticed him. He yielded to help her preserve her sanity and health. His shower may not have warded off HIV, but it was a "ritualistic act of cleansing and strengthening", says Thornton.

Uganda's HIV story offers a more pleasant contrast to South Africa's slow response. South Africa heads global HIV prevalence tables and has little to show in its fight against the virus. Uganda reduced HIV prevalence by two-thirds during the 1990s. Causes continue to be debated. Thornton points to the nature of Ugandan sexual networks. The South African sexual web is similar to information-technology networks with built-in redundancy: when one

transmission node in the Internet breaks down, others take over and information continues to flow. The Ugandan network, Thornton believes, is more fragile. Government and community initiatives knocked out key transmission hubs and the whole network collapsed.

Uganda, even with the recent upsurge in infections, can justifiably look back with pride at its track record of HIV management. 'More of the same', with some fine-tuning as the epidemic evolves, is an adequate prescription. South Africa can learn much from Uganda's success.

What should the rest of the world do? Estimates of annual HIV infections have moved dramatically up and down in many countries in recent years, not because of actual changes in the number of new cases, but owing to refinements in the techniques used to measure them. However, the HIV epidemic seems to have peaked years ago in the United States and parts of Asia. The dreaded generalized epidemic of the South African kind has not, and probably will not, come to pass in populous China and India. HIV/AIDS prevention advocates, especially those who work for the Joint United Nations Programme on HIV/AIDS, are being cast as villains who are diverting resources away from other medical problems, such as diarrhoea

B. STIRTON/GETTY IMAGES



and pneumonia, that continue to kill millions. Meanwhile, even if global HIV transmission comes to a complete halt tomorrow, bills for the treatment of those who are currently infected will keep coming in and will need to be paid for decades by governments, donors, family members or patients themselves. Prevention efforts continue to make economic and humanitarian sense: an infection prevented today equals a lifetime of medical costs saved.

*Unimagined Community* shows how “social

butterflies” and “open skies” are equally important for the transmission of HIV. The transmission networks are sustained by highly mobile infected people who change their sexual partners frequently. Thornton makes a strong case for uncovering the social factors that power these networks, and for developing new prevention efforts to counter them. Sex is a social act, not just a behaviour, so it follows that social interventions are necessary to curb the flow of infection.

The failure of ABC messages — “Abstain, Be faithful, use Condoms if you must graze” — to interrupt HIV transmission must be viewed in the light of Thornton’s study. Sexual networks need to be understood and targeted alongside individual behaviour change. It is time to recruit anthropologists with the training and experience to carry out a professional analysis. ■

**Karunesh Tuli** is a public-health consultant based in Pasadena, California.

e-mail: karunesh tuli@hotmail.com

BUBBLES PHOTOLIBRARY/ALAMY



Patients must have confidence in vaccines to maintain high levels of protection against disease.

## Injecting trust into vaccines

**Autism's False Prophets: Bad Science, Risky Medicine, and the Search for a Cure**  
by Paul A. Offit

Columbia University Press: 2008. 328 pp.  
\$24.95, £14.95

Paul Offit’s distinguished academic credentials and long-standing advocacy for vaccines in the United States provide the weight behind this forceful book. *Autism's False Prophets* focuses on the people and events in that country that were central to the claimed link between vaccination and autism. Written with passion, authority, bluntness and literary skill, it largely lives up to the back cover’s promise of a ‘page-turner’.

The text is rich in heroes and villains. The villains include litigious parents, publicity-seeking journalists, politicians, lawyers and environmental activists, lobbyists and expert witnesses. An assortment of quacks, zealots and incompetents, frequently from within the medical or allied professions, complete

the roll-call of ‘false prophets’. No wonder the public struggles to use good science as the sole arbiter for rational behaviour.

Offit does not underestimate the emotional and financial strains on parents whose children have autism, their compulsion to apportion responsibility for presumed damage, or their rich and positive experiences with their autistic offspring. He is sympathetic to parents who, impatient with the “glacial pace of medical research”, all too often succumb to fashionable cures that fail to deliver. He dismisses the 300 or so US physicians who practice alternative and sometimes damaging ‘remedies’ for autism as “a cottage industry of false hope”.

Two chapters cover the measles, mumps and rubella (MMR) vaccine controversy in the United Kingdom. In 1998, physician Andrew Wakefield published a highly flawed study in *The Lancet* proposing a ‘link’ between the MMR vaccine and autism. At a preceding press conference, he advocated the separation of MMR into three

vaccines until the issue of safety was ‘resolved’. Offit lays bare the weaknesses of Wakefield’s discredited assertions and the questionable ethical practices associated with his work as a physician. Offit also covers the extensive, often uncritical reporting of Wakefield’s view by the UK press, which collectively promoted the unwarranted public anxiety still responsible for the dangerously diminished uptake of the MMR vaccine in the United Kingdom.

Later chapters tell the tale of the mercury-containing compound thimerosal, used since the 1940s as an effective, convenient vaccine preservative. By the late 1990s, some vaccine scientists in the United States were calling for the precautionary abandonment of thimerosal. They feared the rare possibility of subtle neurological and psychological effects from the preservative, although evidence is negligible. In 1999, the US vaccine authorities announced the removal of thimerosal from vaccines, using tortuous sentiments to reassure: “The current levels of thimerosal will not hurt children, but reducing those levels will make safe vaccines even safer.” In a chapter entitled ‘Mercury Rising’, Offit vividly describes how such weasel words opened the floodgates of public concern. Here, he misses an opportunity for international comparisons of scientific and public attitudes — thimerosal anxiety was mainly a US fixation, even though the same preservative was used in the United Kingdom and other countries.

Public belief that vaccines cause autism soon escalated, fuelled by environmentalists, lawyers, politicians and opportunistic scientists publishing in journals of mixed repute. The chapter ‘Mercury Falling’ describes how the accumulating scientific evidence, from more than 200 epidemiological and other analyses, led scientists to refute the notion that thimerosal causes autism. A major factor was that even after the abandonment of thimerosal, rates of autism continued to increase. Over time, the preservative was exonerated to the satisfaction of most critics.

In bemoaning today’s withdrawal of trust



and pneumonia, that continue to kill millions. Meanwhile, even if global HIV transmission comes to a complete halt tomorrow, bills for the treatment of those who are currently infected will keep coming in and will need to be paid for decades by governments, donors, family members or patients themselves. Prevention efforts continue to make economic and humanitarian sense: an infection prevented today equals a lifetime of medical costs saved.

*Unimagined Community* shows how “social

butterflies” and “open skies” are equally important for the transmission of HIV. The transmission networks are sustained by highly mobile infected people who change their sexual partners frequently. Thornton makes a strong case for uncovering the social factors that power these networks, and for developing new prevention efforts to counter them. Sex is a social act, not just a behaviour, so it follows that social interventions are necessary to curb the flow of infection.

The failure of ABC messages — “Abstain, Be faithful, use Condoms if you must graze” — to interrupt HIV transmission must be viewed in the light of Thornton’s study. Sexual networks need to be understood and targeted alongside individual behaviour change. It is time to recruit anthropologists with the training and experience to carry out a professional analysis. ■

**Karunesh Tuli** is a public-health consultant based in Pasadena, California.

e-mail: karunesh tuli@hotmail.com

BUBBLES PHOTOLIBRARY/ALAMY



Patients must have confidence in vaccines to maintain high levels of protection against disease.

## Injecting trust into vaccines

**Autism's False Prophets: Bad Science, Risky Medicine, and the Search for a Cure**  
by Paul A. Offit

Columbia University Press: 2008. 328 pp.  
\$24.95, £14.95

Paul Offit’s distinguished academic credentials and long-standing advocacy for vaccines in the United States provide the weight behind this forceful book. *Autism's False Prophets* focuses on the people and events in that country that were central to the claimed link between vaccination and autism. Written with passion, authority, bluntness and literary skill, it largely lives up to the back cover’s promise of a ‘page-turner’.

The text is rich in heroes and villains. The villains include litigious parents, publicity-seeking journalists, politicians, lawyers and environmental activists, lobbyists and expert witnesses. An assortment of quacks, zealots and incompetents, frequently from within the medical or allied professions, complete

the roll-call of ‘false prophets’. No wonder the public struggles to use good science as the sole arbiter for rational behaviour.

Offit does not underestimate the emotional and financial strains on parents whose children have autism, their compulsion to apportion responsibility for presumed damage, or their rich and positive experiences with their autistic offspring. He is sympathetic to parents who, impatient with the “glacial pace of medical research”, all too often succumb to fashionable cures that fail to deliver. He dismisses the 300 or so US physicians who practice alternative and sometimes damaging ‘remedies’ for autism as “a cottage industry of false hope”.

Two chapters cover the measles, mumps and rubella (MMR) vaccine controversy in the United Kingdom. In 1998, physician Andrew Wakefield published a highly flawed study in *The Lancet* proposing a ‘link’ between the MMR vaccine and autism. At a preceding press conference, he advocated the separation of MMR into three

vaccines until the issue of safety was ‘resolved’. Offit lays bare the weaknesses of Wakefield’s discredited assertions and the questionable ethical practices associated with his work as a physician. Offit also covers the extensive, often uncritical reporting of Wakefield’s view by the UK press, which collectively promoted the unwarranted public anxiety still responsible for the dangerously diminished uptake of the MMR vaccine in the United Kingdom.

Later chapters tell the tale of the mercury-containing compound thimerosal, used since the 1940s as an effective, convenient vaccine preservative. By the late 1990s, some vaccine scientists in the United States were calling for the precautionary abandonment of thimerosal. They feared the rare possibility of subtle neurological and psychological effects from the preservative, although evidence is negligible. In 1999, the US vaccine authorities announced the removal of thimerosal from vaccines, using tortuous sentiments to reassure: “The current levels of thimerosal will not hurt children, but reducing those levels will make safe vaccines even safer.” In a chapter entitled ‘Mercury Rising’, Offit vividly describes how such weasel words opened the floodgates of public concern. Here, he misses an opportunity for international comparisons of scientific and public attitudes — thimerosal anxiety was mainly a US fixation, even though the same preservative was used in the United Kingdom and other countries.

Public belief that vaccines cause autism soon escalated, fuelled by environmentalists, lawyers, politicians and opportunistic scientists publishing in journals of mixed repute. The chapter ‘Mercury Falling’ describes how the accumulating scientific evidence, from more than 200 epidemiological and other analyses, led scientists to refute the notion that thimerosal causes autism. A major factor was that even after the abandonment of thimerosal, rates of autism continued to increase. Over time, the preservative was exonerated to the satisfaction of most critics.

In bemoaning today’s withdrawal of trust

from medical professionals, Offit risks overegging his case for science. Precaution is sometimes a wise basis for policy-making. Vaccine damage through unforeseen effects does occasionally occur and may be severe. As Offit reminds us, one in every 100,000 patients suffered paralysis as a consequence of the US swine flu vaccination in 1976. Such rare episodes do not detract from the transformational effect of vaccines. But they remind us that not all concerns about vaccines can be dismissed as irrational.

Offit reports the personal testimonies of caring parents to good effect, but despairs in equal measure about their lack of scientific credibility and strong emotional appeal. One celebrity mother of an autistic child said, for example, "My science is Evan, and he's at home." Offit is anxious about patients actively participating in their own medical care using the Internet. For him, scientifically sound information online is drowned by poorly substantiated opinion.

Yet Offit's arguments falter when he uses vaccine controversies as his indicator of the presumed ignorance of science by the public. Communicating risk is notoriously tricky. Vaccine uptake needs to be sufficiently high to ensure protection of the population through 'herd immunity' — a persuasive task made all the more difficult when individual patients are urged to accept responsibility for their health and embrace the principle of choosing a hospital or consultant.

*Autism's False Prophets* encapsulates the fanciful belief among scientists, not supported by those who research science communication, that understanding the science will inevitably tip the scales of public opinion. This stance was demonstrated earlier this year in a letter to *The Guardian* newspaper from David Salisbury, director of immunization at the UK Department of Health. It draws attention to the ongoing UK public information campaign for adopting a vaccine intended to protect young women from human papilloma virus, the cause of most cervical cancer. The campaign, Salisbury wrote, aims to ensure that "parents and young women have all the information they need to consent to this important vaccine". Absent in such sentiments is a belief in the necessity for dialogue with the public, entirely different in tone and purpose from the mistaken compulsion of some well-intentioned vaccine enthusiasts to inject hard facts into empty vessels.

**Jeff Thomas** teaches science communication and health sciences in the Department of Life Sciences at The Open University, Milton Keynes, MK7 6AA, UK. He is co-editor of *Practising Science Communication in the Information Age*. e-mail: j.n.thomas@open.ac.uk



Dark materials: Karel Nel's art explores the seen and unseen Universe.

ART FIRST, LONDON

## Q&A: Creations from the cosmos

Artist **Karel Nel** works with astronomers from COSMOS, the global Cosmic Evolution Survey that is mapping galaxies and dark matter. Now exhibiting his work in London, he tells *Nature* how his view of the Universe has changed.

### How did you get involved in astronomy?

Artists and scientists have questioned the nature of reality for centuries. Feeling the need to grasp contemporary scientific paradigms, I had worked for decades at the interface between these disciplines when I met Nick Scoville, leader of the COSMOS project. He invited me to be its resident artist. It was a very steep learning curve: I felt like an ant being taught compound interest by an economist.

### What does your work convey?

My art investigates seen and unseen worlds. COSMOS looks back in deep time at patterns of galaxy formation and large-scale structures, and from these, attempts to understand invisible dark energy and dark matter. I use metaphorical means to grasp these abstract ideas, as scientists often do. In my 20 exhibited works, I use mixed media including 540-million-year-old black carboniferous dust and white primordial salts from the oceans to present shimmering images of galaxies that emitted their light millions of years ago.

### Have the scientists influenced your ideas?

Yes — representing the Universe is not like painting a traditional landscape; there are invisible as well as visible aspects to

convey. Scientists have developed codes to deal with cosmic phenomena, and my work captures the unstable nature of our perceptions of this distant, unknown terrain. In one piece, dotted lines echo the grids found in astronomy textbooks, but also refer to invisible characters, as used in comic books. In another, I evoke the 'blind spots' of telescopes with amorphous dark shapes.

### Did you influence the astronomers?

Many of the astronomers focus on incremental, detailed information, so my broad outsider's perception rekindled the extraordinariness of their endeavour.

### What was it like, visiting observatories?

At the summit of Mauna Kea, Hawaii, some of the world's most powerful telescopes are trained on the powder-black darkness, looking at complexity and eternity. It is awesome and desolate. Even the scientists fall silent in the face of that.

Interview by **Jennifer Rohn**, a researcher at University College London, Gower Street, London WC1E 6BT, UK, and editor of [www.lablit.com](http://www.lablit.com). e-mail: [jenny@lablit.com](mailto:jenny@lablit.com)

### The Brilliance of Darkness

Art First, 9 Cork Street, London W1S 3LL  
Until 9 October 2008.

## Enhance your life with *Nature* debates

*Nature* has picked two panels of experts in science, policy and ethics to debate research that is improving mental and physical abilities.

### Enhancing the Brain: 13 October 2008

From intelligence to emotional tolerance, sleep requirements and memory power, how are developments in neuroscience affecting the individual and society? Panel: Barbara Sahakian, John Harris and Nick Bostrom.

### Enhancing the Body: 10 November 2008

Studies of the human body are focusing on aspects such as speed, strength and healing or the tolerance of pain. But how will this science enhance the human body? Panel: Aubrey de Grey, Andy Miah and Kevin Warwick.

*Nature* debates will be held at 7pm at Kings Place, 90 York Way, London N1 9AG, UK. See <http://tinyurl.com/4zrh8v> for more details.



from medical professionals, Offit risks over-egging his case for science. Precaution is sometimes a wise basis for policy-making. Vaccine damage through unforeseen effects does occasionally occur and may be severe. As Offit reminds us, one in every 100,000 patients suffered paralysis as a consequence of the US swine flu vaccination in 1976. Such rare episodes do not detract from the transformational effect of vaccines. But they remind us that not all concerns about vaccines can be dismissed as irrational.

Offit reports the personal testimonies of caring parents to good effect, but despairs in equal measure about their lack of scientific credibility and strong emotional appeal. One celebrity mother of an autistic child said, for example, "My science is Evan, and he's at home." Offit is anxious about patients actively participating in their own medical care using the Internet. For him, scientifically sound information online is drowned by poorly substantiated opinion.

Yet Offit's arguments falter when he uses vaccine controversies as his indicator of the presumed ignorance of science by the public. Communicating risk is notoriously tricky. Vaccine uptake needs to be sufficiently high to ensure protection of the population through 'herd immunity' — a persuasive task made all the more difficult when individual patients are urged to accept responsibility for their health and embrace the principle of choosing a hospital or consultant.

*Autism's False Prophets* encapsulates the fanciful belief among scientists, not supported by those who research science communication, that understanding the science will inevitably tip the scales of public opinion. This stance was demonstrated earlier this year in a letter to *The Guardian* newspaper from David Salisbury, director of immunization at the UK Department of Health. It draws attention to the ongoing UK public information campaign for adopting a vaccine intended to protect young women from human papilloma virus, the cause of most cervical cancer. The campaign, Salisbury wrote, aims to ensure that "parents and young women have all the information they need to consent to this important vaccine". Absent in such sentiments is a belief in the necessity for dialogue with the public, entirely different in tone and purpose from the mistaken compulsion of some well-intentioned vaccine enthusiasts to inject hard facts into empty vessels.

**Jeff Thomas** teaches science communication and health sciences in the Department of Life Sciences at The Open University, Milton Keynes, MK7 6AA, UK. He is co-editor of *Practising Science Communication in the Information Age*. e-mail: j.n.thomas@open.ac.uk



Dark materials: Karel Nel's art explores the seen and unseen Universe.

ART FIRST, LONDON

## Q&A: Creations from the cosmos

Artist **Karel Nel** works with astronomers from COSMOS, the global Cosmic Evolution Survey that is mapping galaxies and dark matter. Now exhibiting his work in London, he tells *Nature* how his view of the Universe has changed.

### How did you get involved in astronomy?

Artists and scientists have questioned the nature of reality for centuries. Feeling the need to grasp contemporary scientific paradigms, I had worked for decades at the interface between these disciplines when I met Nick Scoville, leader of the COSMOS project. He invited me to be its resident artist. It was a very steep learning curve: I felt like an ant being taught compound interest by an economist.

### What does your work convey?

My art investigates seen and unseen worlds. COSMOS looks back in deep time at patterns of galaxy formation and large-scale structures, and from these, attempts to understand invisible dark energy and dark matter. I use metaphorical means to grasp these abstract ideas, as scientists often do. In my 20 exhibited works, I use mixed media including 540-million-year-old black carboniferous dust and white primordial salts from the oceans to present shimmering images of galaxies that emitted their light millions of years ago.

### Have the scientists influenced your ideas?

Yes — representing the Universe is not like painting a traditional landscape; there are invisible as well as visible aspects to

convey. Scientists have developed codes to deal with cosmic phenomena, and my work captures the unstable nature of our perceptions of this distant, unknown terrain. In one piece, dotted lines echo the grids found in astronomy textbooks, but also refer to invisible characters, as used in comic books. In another, I evoke the 'blind spots' of telescopes with amorphous dark shapes.

### Did you influence the astronomers?

Many of the astronomers focus on incremental, detailed information, so my broad outsider's perception rekindled the extraordinariness of their endeavour.

### What was it like, visiting observatories?

At the summit of Mauna Kea, Hawaii, some of the world's most powerful telescopes are trained on the powder-black darkness, looking at complexity and eternity. It is awesome and desolate. Even the scientists fall silent in the face of that.

Interview by **Jennifer Rohn**, a researcher at University College London, Gower Street, London WC1E 6BT, UK, and editor of [www.lablit.com](http://www.lablit.com). e-mail: [jenny@lablit.com](mailto:jenny@lablit.com)

### The Brilliance of Darkness

Art First, 9 Cork Street, London W1S 3LL  
Until 9 October 2008.

## Enhance your life with *Nature* debates

*Nature* has picked two panels of experts in science, policy and ethics to debate research that is improving mental and physical abilities.

### Enhancing the Brain: 13 October 2008

From intelligence to emotional tolerance, sleep requirements and memory power, how are developments in neuroscience affecting the individual and society? Panel: Barbara Sahakian, John Harris and Nick Bostrom.

### Enhancing the Body: 10 November 2008

Studies of the human body are focusing on aspects such as speed, strength and healing or the tolerance of pain. But how will this science enhance the human body? Panel: Aubrey de Grey, Andy Miah and Kevin Warwick.

*Nature* debates will be held at 7pm at Kings Place, 90 York Way, London N1 9AG, UK. See <http://tinyurl.com/4zrh8v> for more details.

# Beyond the greenhouse

Botanic gardens are using good garden design to attract and educate the public. **Mike Maunder** explains how they can thrive both as businesses and as institutions of learning.

The world's 2,500 botanic gardens are broadening their conservation purpose and embracing their cultural identity. Rallying to help overcome the challenges of the global biodiversity crisis, herbaria and seed banks have been rejuvenated. Meanwhile, an equally profound change has occurred in the public face of the botanic garden.

Botanic gardens retain their core functions of collecting and exhibiting plants for scientific and educational purposes. During the past 30 years, botany research at universities has declined and botanic gardens have become key to documenting plant diversity and promoting plant conservation, many as important players in the United Nations' Global Strategy for Plant Conservation ([www.bgci.org/worldwide/gspc](http://www.bgci.org/worldwide/gspc)). As the extinction spasm accelerates, the skills of such institutions in managing threatened species will be vital. The accumulated plant collections, living and preserved, record what we are losing and offer resources for what we choose to restore. These collections are of immense scientific and cultural value, and their viability presents a financial and political challenge.

Botanic gardens are becoming sophisticated business entities and increasingly depend upon the financial patronage of the public. This is a profound change from the days when many gardens were supported by government funds and visitors were seen as an awkward impediment, coming between curators and horticultural perfection. During the past 20 years, botanic gardens have found a vibrant sense of mission as translators of environmental science to millions of people of all ages. They recognize that their future depends on conveying the importance of their work and establishing strong links with their local community.

## Design for delight

This new social role has changed the design of botanic gardens dramatically. Compare the quiet, walled enclaves of the early Italian Renaissance botanic gardens with the space-age domes and invigorating botanical cabaret of the Eden Project in Cornwall, UK. The defining features of old botanic gardens have largely disappeared, such as the awkward opening times, miserable cafes and lack of any institutional interpretation. Gardens are now creating exhibits that inform and inspire: in the words of the philosopher and theologian Thomas Aquinas, "you change people by delight".



The Royal Botanic Gardens at Cranbourne, Australia, convey the nature of the continent's 'Red Centre'.

This evolution is particularly notable in the latest desert botanic gardens. Here, an appreciation of local floras, declining natural resources and ethnobotanical heritage are influencing both mission and design. The Arizona Sonora Desert Museum was established as an early model of this in 1952 — a mix of botanic garden, zoo and museum, with a clear geographical and ecological identity. It has inspired a new generation of desert gardens. The Royal Botanic Garden near Amman in Jordan showcases the country's natural habitats and wild species; similarly, the new Oman Botanic Garden near Muscat exhibits native plants, ecosystems and ethnobotanical heritage. The Al Ain Wildlife Park in Abu Dhabi, set to open in 2010, will demonstrate the biodiversity and cultures of the world's desert ecosystems and reconnect the people of the United Arab Emirates with their desert heritage. In Australia, the Alice Springs Desert Park interprets the local ecology, culture and landscape. The spectacular new Australian Garden at the Royal Botanic Gardens, Cranbourne, near Melbourne, abstracts those same landscapes into a contemporary garden design, acting as envoy for the culture and biology of the 'Red Centre'.

Strong collaboration between scientists and artists coupled with a locally focused mission

defines the character of some new gardens. In the late 1990s, Mexican artists Francisco Toledo and Luis Zárate worked with anthropologist Alejandro de Ávila to create dramatic landscapes incorporating pre-Hispanic motifs in the Ethnobotanical Garden of Oaxaca, Mexico, celebrating the natural heritage of Oaxaca state. A similar approach can be seen in the First Nations Garden at the Montreal Botanical Garden, a cultural meeting place where the ethnobotany of Canada's Inuit and Native American cultures is exhibited and interpreted through habitat exhibits and storytelling.

China, with its traditions of academic botany and love of plants, has seen an extraordinary growth in botanic gardens in recent decades. From just 34 in 1960, China now boasts more than 160 gardens, housing some of the world's largest and richest cultivated plant collections. New gardens are being created, including the Shanghai Chenshan Botanical Garden, and existing gardens, such as the Hangzhou Botanical Garden, are being renovated as spectacular venues for the public display of botanical diversity. Similarly, the restored Hengchun Tropical Botanical Gardens in Taiwan reflect the geomorphological, ecological and cultural evolution of Taiwan and celebrate the beauty of the site's tropical-forest habitat.

This new emphasis on gardens as public

P. HYATT/TAYLOR CULTURE LETHLEAN WITH P. THOMPSON



spectacles has arisen largely because of the recent interchange of designers between botanic gardens, museums and zoos. Landscape architecture studios — such as Jon Coe in Australia, or Jones and Jones, Field Operations and Portico in the United States — are applying techniques honed in zoos and museums to create innovative botanic-garden exhibits. Landscape architects experienced in designing for private clients are also creating astonishing botanic-garden landscapes; examples include the work of Raymond Jungles and Made Wijaya at the Naples Botanic Garden in Florida and Luis Vallejo's design for the Oman Botanic Garden.

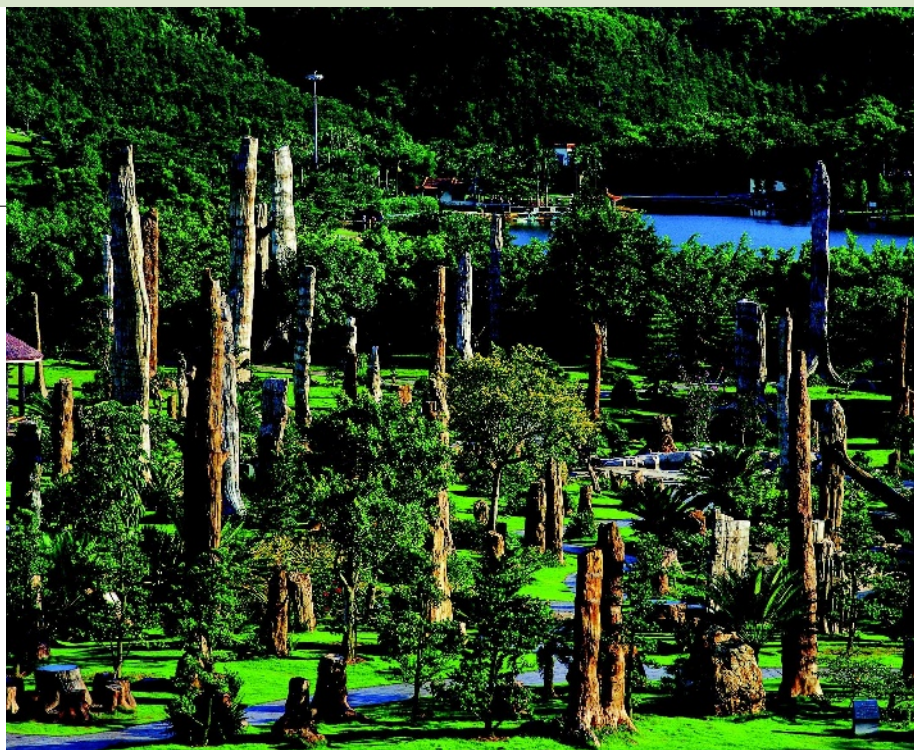
Within the botanic-garden community, there is a palpable sense of confidence. New exhibition spaces showcase their vital scientific work and collections that had historically remained hidden from the visitor. For example, the public galleries of the Millennium Seed Bank at the Royal Botanic Gardens, Kew, UK, convey their role in conserving wild plant resources, and the Pfizer Plant Research Laboratory at the New York Botanic Garden invites visitors to see the workings of a major taxonomic research institute. "Where better than a botanic garden for people to understand the rapidly changing world, to grasp how to change their behaviour and to mobilize action for a sustainable future?" says Stephen Blackmore, regius keeper of the Royal Botanic Gardens in Edinburgh, UK.

Many botanic gardens and their buildings demonstrate sustainable maintenance and construction techniques. The new visitor centre at the Queens Botanical Garden, described as the greenest building in New York City, incorporates the use of solar and geothermal energy, waste-water recycling and composting toilets. As Queens' executive director Susan Lacerte says: "We are an environmental organization; if we are not going to build green, who is?"

B. FOSTER/EDEN PROJECT



The UK Eden Project promotes sustainability.



LIU BI-JIAN

A forest of fossilized and living trees at Shenzhen Fairy Lake Botanical Garden in Guangdong, China.

### Crowd pleasing

Amid all this vigour, botanic gardens face a challenge: the temptation to reduce investment in plant collections — a core resource — and emphasize exhibits that generate income. The management of collections and the business of displaying them have historically been separate affairs, so it is encouraging to see important collections now being used in innovative ways. The new Rhizotron and Xstrata Treetop Walkway at the Royal Botanic Garden at Kew have brought life and a story to a traditional arboretum. Plant collections are increasingly packaged as habitats, either generically as desert or rainforest exhibitions, or specifically linked to conservation projects. For instance, the Lin Loughheed Spiny Forest exhibit at Fairchild Tropical Botanic Garden, Florida, is directly linked to supporting field conservation in Madagascar.

Art has always had a place in botanic gardens, and it is being used with gusto for interpretation and marketing. Previously hidden treasures are being given new homes, as with the botanical drawings at the Shirley Sherwood Gallery of Botanical Art at Kew. Some historical collections and landscapes are being given a contemporary interpretation. One example is sculptor Mark Dion's interpretation of botanist William Bartram's explorations of the southern United States in 1773, on show at Bartram's Garden in Philadelphia, Pennsylvania. Environmental degradation is highlighted in the provocative *Hard Rain* exhibit, which is touring various botanic gardens worldwide and combines the lyrics of Bob Dylan with Mark Edwards's apocalyptic photographs. Similarly, in the Garden of Extinction, designed by Willem Boshoff and installed at Kirstenbosch Botanic Garden in South Africa, rows of memorials represent 15,000 threatened or extinct plant species,

with disturbing echoes of the First World War cemeteries in Belgium and France.

We have entered the age of blockbuster art in botanic gardens. Roy Lichtenstein's work has been exhibited at Fairchild Tropical Botanic Garden, Niki de St Phalle's art is displayed at Missouri Botanical Garden, St Louis, and Dale Chihuly has exhibited at Kew and the Atlanta Botanical Garden in Georgia. When successful, these shows have a steroidal effect on visitor attendance and membership.

The showman P. T. Barnum said "every crowd has a silver lining", and botanic gardens are using special events to attract new visitors and interpret big environmental issues. Events such as the Chocolate and Mango Festivals at Fairchild, for example, not only generate revenue and membership but also connect the public with the agricultural heritage of the tropics and the challenges of sustainable agriculture.

A deep satisfaction can come from experiencing the diversity and abundance of a good botanic garden, whether it contains tropical orchids and palms, an astonishing prairie reconstruction or a seasonal display of apples or chilli peppers. The challenge is to use that emotional response to change people's behaviour. Botanic gardens need to create landscapes and gardens that are attractive, vibrant and welcoming — otherwise, they will fail as businesses, and thence in their fundamental mission to inform and influence. ■

**Mike Maunder** is executive director of the Fairchild Tropical Botanic Garden, Coral Gables, Miami, Florida 33156-4233, USA, and a faculty member of the Department of Biological Sciences, Florida International University. e-mail: [mmaunder@fairchildgarden.org](mailto:mmaunder@fairchildgarden.org)

More garden images at <http://tinyurl.com/5yv2q8>.

## ESSAY

## Beijing 1987: China's coming-out party

Two decades ago, Deng Xiaoping welcomed nations to an international meeting in Beijing. **Mohamed Hassan** recalls how China's leaders set out their plans for the nation to rejoin the world's scientific elite.

The memory is still fresh. On 12 September 1987, I flew into a Beijing airport that comprised a few badly maintained structures, no larger than a small provincial airfield. All around was peeling paint and dim lighting, no signs, no assistance and an anxious wait for luggage.

The drive from the airport was another eye-opener. The road we took was poorly paved. Street lighting was sporadic. A smattering of cars, mainly Japanese-made, had to make room for bicycles, wheelbarrows and the occasional donkey cart. Beijing was a busy, even frenetic place. It was certainly well cared for. But it was impoverished.

The Beijing from the window of my car was the image of China in the West. Yet, over the next five days, the idea of China as a poverty-stricken backwater, unable to harness new ways of learning and ultimately modernity, would be tested and changed forever.

As I and the other delegates were about to discover, China's most prominent officials and scientists all agreed that science and technology would be a key element in the country's plans to re-emerge as a global power. As one of my colleagues, MGK Menon, former chief scientific adviser to India's late prime minister Indira Gandhi, recalls today: "the state of science in China was a welcome surprise but largely because so little was known about it."

### Drawn together

I was then starting my career at TWAS, then known as the Third World Academy of Sciences and the brainchild of the Pakistani theoretical physicist Abdus Salam. Salam had lured me to the academy from the University of Khartoum in Sudan, where I was a professor of mathematics, specializing in modelling the formation and the movement of desert dust and sand.

Salam knew China well; his contacts were at the highest levels. He also knew that China's leadership wanted both to open up to the West and to have a leading role for the developing nations of the South. So it was no surprise for me to learn that our second general conference would in effect become China's coming-out party — an opportunity for the country to show itself off

to the world (both North and South) and a chance for enterprising scientists and policy-makers from the developing countries to learn from — and do business with — Beijing.

Salam had other reasons for wanting to work with China. He was a candidate for the job of running UNESCO, the United Nations Educational, Scientific and Cultural Organization. As a Nobel prize-winning professor at Imperial College London, Salam had little difficulty accessing the levers of power in the Anglophone world. But to secure the UNESCO post he needed the backing of countries elsewhere. Getting a nod from a permanent member of the UN Security Council, such as China, would certainly help.

### China anew

The opening session took place in the Great Hall of the People two days later on 14 September. There were more than 150 participants from 50 countries, joined by an equal number from China. Given that this was a rare example of China putting its science on display, interest from the Western scientific community was high. International organizations represented in Beijing included the then International Council of Scientific Unions (ICSU), headed by Julia Marton-LaFère, the International Foundation for Science (IFS), the American Association for the Advancement of Science (AAAS) and the Science Council of Japan.

The Chinese Academy of Sciences, an institution that is key to understanding science in modern China, was our host. Whereas, in Western countries, government ministries, research funding bodies and scientists operate (mostly) independently of each other. In China (as in Russia), all three of these functions are rolled into a single state-run and state-funded academy of sciences. The Chinese academy, which was established just two months after the Communist takeover in 1949, had grown substantially in size and impact over the next two decades. However, the assault on intellectuals during the 10-year Cultural Revolution severely put back the academy's work.

The scars of this period had clearly not yet

healed for many of the Chinese speakers at the 1987 conference. In his opening address, Zhou Guangzhao, then president of the Chinese Academy of Sciences, described the Cultural Revolution as a "disaster" for China. Similarly, Zhao Ziyang, general secretary of the Communist Party, said that bullying and humiliation had characterized the Cultural Revolution.

China's leader, Deng Xiaoping, later told Salam that if China did eventually join the ranks of developed nations, it would never forget its history and where it came from: that it would always see itself as a developing country first and foremost. But that said, out of the embers of the Cultural Revolution, it harboured larger ambitions.

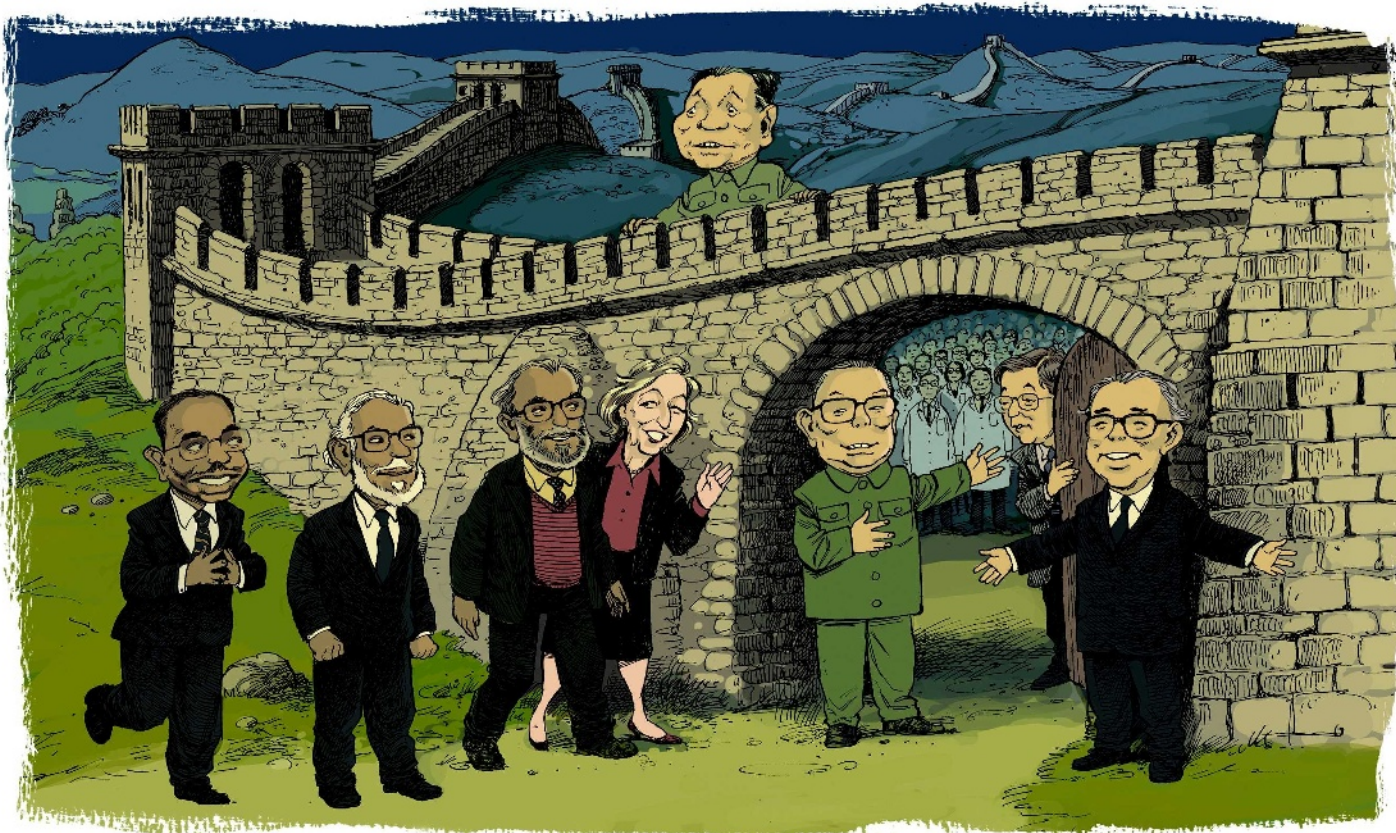
Too often, the main message coming out of conferences on science and technology in developing countries was (and still is) one of inadequate salaries, low standards of research and development, poor working conditions and either political apathy or excessive political interference. Beijing 1987 could have been the same. However, our hosts were determined to show that they would not be defined by the past.

We learnt from the opening talks that, in 1949, China, with a population of nearly 550 million, had just 50,000 people recorded as working in science and technology in some 30 scientific institutes. But by 1985, the country had more than 10 million people working in almost all fields of science and technology, including 300,000 active researchers. The Chinese Academy of Sciences alone had 80,000 scientists working in more than 120 research institutes. The academy system was complemented by a rapidly growing system of higher education consisting of 2 million students in more than 1,000 universities.

We also learnt that China's scientists had been busy pushing the envelope in a diverse set of research fields. Among other achievements, they had synthesized bovine insulin, carried out research into low-temperature, high-conducting materials, launched 19 satellites and developed an extensive nationwide remote-sensing system. In agriculture, China had eliminated the wheat-rust fungus and created hybrid rice varieties, allowing improved crop yields. In health care, vaccines had led to significant reductions in the incidence of diseases such as diphtheria, scarlet fever and polio.

**"The state of science in China was a welcome surprise but largely because so little was known about it."**





D. PARKINS

Listening to the presentations and observing China's scientists, it became clear that this kind of progress was different from that taking place in other Communist countries — especially the Soviet Union. Unlike former Soviet Russia, for example, China had no problems in acknowledging that it would need access to Western science and technology to innovate. At the time of the conference, an estimated 100,000 students from China had studied or worked abroad (mostly in Western countries) and then returned home to help build their nation's infrastructure. This number is now in excess of one million.

Finally, we learnt that China, which had begun to open the door of scientific exchange with both developed and developing countries, was eager to collaborate with partners in a broad range of fields. To facilitate this, Chinese officials announced that they would be hosting a TWAS–China field office inside the Chinese Academy of Sciences.

This was a truly transformational moment. China was telling us that it was no longer alone in science, that it would welcome opportunities to become part of the international scientific community and that it was eager to join with others both to enhance its own capabilities and to help address critical global challenges in both science and society. Biotechnology,

materials science, particle physics and space science were about to get a new major partner, and both China and the rest of the world would benefit as a result.

### Back to the future

I returned to China in October 2003 to celebrate TWAS's 20th anniversary. My journey began in Trieste and ended at Beijing's international airport, just as it had 15 years earlier. The opening ceremony, just as before, was held in the Great Hall of the People.

I may have followed the same roads to the same places in Beijing, but so much else had changed. The airport was a gleaming new state-of-the-art facility and one of the biggest and busiest airports in the world. The tree-lined avenues that led from the airport to the city centre were jammed with automobiles and taxis, and framed by sleek apartment houses, office buildings and commercial establishments. Enormous holes in the ground, shadowed by large construction cranes overhead, foretold of larger structures to come, including the venues for the 2008 Olympics.

Hu Jintao, China's president, spoke at the opening ceremony in front of 3,000 people. The audience consisted not just of scientists

and political leaders from around the world, but also of many young Chinese students who filled the outer reaches of the great hall. Earlier that day, Hu had congratulated China's first astronaut, Yang Liwei, who had just returned home after completing a 21-hour, 14-orbit journey around Earth. The space flight, President Hu observed, constituted "another important step for the Chinese people scaling to the summit of world science and technology".

As I listened to President Hu and peered at the audience beyond the podium, I could not help but think that one of the most important first steps

in China's rebirth as an international scientific powerhouse took place in the same hall in 1987. The promise of modern science and technology in China that I had first glimpsed 15 years ago had been fulfilled. ■

Mohamed Hassan is the executive director of TWAS, the academy of sciences for the developing world in Trieste, Italy. He was a participant at the academy's second general conference, held in Beijing from 14–18 September 1987. e-mail: mhasan@twas.org

**"Our hosts were determined to show that they would not be defined by the past."**

For more Meetings that Changed the World, see [www.nature.com/nature/focus/meetings](http://www.nature.com/nature/focus/meetings)



## NEWS &amp; VIEWS

## SENSORY ECOLOGY

## In sight of speciation

Mark Kirkpatrick and Trevor Price

**Adaptation of a fish's eyes to its visual environment can bias females to mate with different males according to their coloration. This sensory preference can contribute to the formation of new species.**

How and why do barriers that prevent mating between species evolve? On page 620 of this issue, Seehausen *et al.*<sup>1</sup> present a rich and eclectic data set that suggests a key role for vision in African cichlid fishes. It has been shown in other fish that natural selection tunes eyes to their visual environment, so that individuals can best see not only what they eat and what eats them, but also members of their own species<sup>2–4</sup>. Seehausen *et al.* carry this story a step further with work on fish in which the males are either red or blue (Fig. 1), and which have genetic variation for visual sensitivity to those colours.

In some populations, females with blue-biased vision seem to mate only with blue males, whereas red-biased females mate only with red males. The inference is that natural selection acting on the visual system contributes to reproductive barriers and the formation of new species. In short, what you see determines what you get, and with whom you get it on. More controversially, the authors suggest that these barriers might arise within a population, and do not, as has previously been thought, require a phase in which red and blue populations evolve in geographical isolation.

The biological and ecological setting for this story is dramatic — the cichlid fish in the Great Lakes of Africa. These fish are the most rapidly speciating organisms on Earth, and this explosion of life has produced a panoply of colour, morphology and behaviour, a sampling of which can be seen at your local pet shop. The fish in Lake Victoria, where the present study<sup>1</sup> was done, show a fantastically high rate of speciation. More than 500 species inhabit the lake. They may have originated just a few hundred thousand years ago<sup>5</sup>, and possibly went through a period of large-scale interbreeding 20,000 years ago<sup>6</sup>.

The lake has diverse visual environments. Starting at the shore and descending along the lake bottom, red becomes increasingly dominant in the ambient visual spectrum. This spectral shift is rapid at some sites and more gradual at others. To study how the fish

have adapted to these conditions, Seehausen *et al.* wanted to know what the fish see. They identified genetic variants (alleles) in one of the opsin genes responsible for tuning the fish's visual sensitivity to different colours. By expressing these genes *in vitro* and measuring the absorption properties of the resulting proteins, they found some variants that are red-biased in their sensitivity and others that are blue-biased. The red-biased variant is typically found in fish living at greater depths than the blue-biased one.

The numbers of males with red, blue and intermediate coloration vary between populations. At sites where the spectral shift is neither very rapid nor very gradual, notably Makobe island, blue males are confined to the shallows and red males to greater depths. At this site, the great majority of

blue males carry the blue-biased opsin variant, whereas most red males carry the red-biased one. The two colour morphs also show differences in other genetic markers, suggesting that they are nascent species. At sites where the spectral shift is rapid, however, the colour forms interbreed, presumably because they encounter each other frequently.

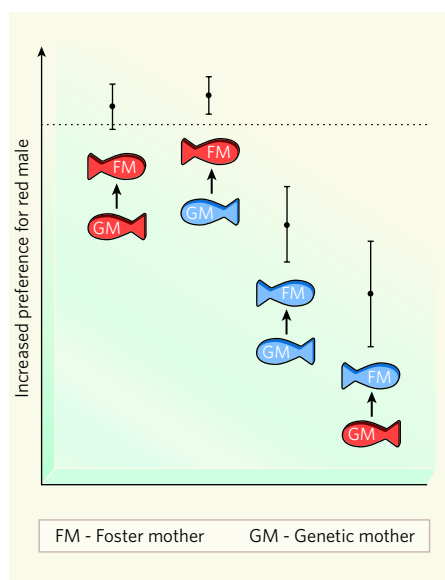
Is beauty just in the eye of the beholder? In mate-choice experiments using fish from controlled crosses, Seehausen *et al.* find that the opsin variant alone does not strongly determine mating preference. Segregation of the colour morphs by depth in the lake must mean that the fish mainly encounter and hence mate with their own kind. It is not difficult to imagine that fish prefer to spend time in habitats in which they see best — that is, visual tuning could generate a type of habitat preference that

O. SEEHAUSEN &amp; I. VANDER SLUIS



**Figure 1 | Seeing red doesn't get you blue.** These cichlid fish are examples of the colour morphs investigated by Seehausen *et al.*<sup>1</sup> in their study of visually determined mating preference.





**Figure 2 | The mother who raised you determines with whom you mate.** In choice tests<sup>7</sup> with mouth-brooding cichlids, female fish raised by foster mothers belonging to different species from their genetic mothers preferred males of that different species. The stylized pictures are of males; females are dull and generally similarly coloured. The preference was measured by the difference in the number of approaches per male display when female fish are given a pairwise choice (with standard errors; dashed line indicates random choice). This evidence suggests that learning at a young age contributes to reproductive isolation in cichlids, in addition to other mechanisms such as the action of natural selection on vision described by Seehausen and colleagues<sup>1</sup>. (Modified from ref. 7.)

contributes to speciation above and beyond its effects on mate choice.

But is even this enough? Other findings point to an additional mechanism that complements reproductive isolation via vision. The females of these remarkable fish brood their eggs in their mouths, then guard the young fry after they hatch. In experiments reported last year, Verzijden and ten Cate<sup>7</sup> swapped eggs between the mouths of red morph and blue morph mothers. Females raised from the experimental broods strongly preferred males from their foster morph over those of their own morph (Fig. 2). As females of the two species look very similar, it is unclear whether the offspring preference is based on colour or some other correlated cue such as odour. Regardless of that, learning at a young age (sexual imprinting) apparently contributes to reproductive isolation in these cichlids, as it does in other groups such as birds<sup>8</sup>. The implication is that assortative mating — the tendency of like to mate with like — can arise whenever male characteristics diverge in response to differences in the environment, which might happen even without divergence in the opsin pigments. It remains to be seen if imprinting, vision and perhaps other mechanisms have been sufficient to generate

new species without geographical isolation.

An intriguing observation mentioned by Seehausen *et al.*<sup>1</sup> is that the red- and blue-biased opsin alleles are evolutionarily much older than the species studied here. Red and blue colour morphs are found in other species of cichlid<sup>9</sup>, suggesting that the colour polymorphism may also be ancient. Perhaps one key to the spectacular species radiation of African cichlids is that they inherited from distant ancestors a trove of genetic variation for sensory systems and male signals, possibly contributed during the inferred episode of interbreeding 20,000 years ago. This variation is entrained again and again in speciation events. To systematists, these events represent independent nodes on the evolutionary tree. From the fish's point of view, however, they are perhaps more like an evolutionary play that is re-enacted, night after night, with the same genetic cast. ■

Mark Kirkpatrick is in the Section of Integrative Biology, University of Texas, Austin, Texas 78712, USA. Trevor Price is in the Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA.  
e-mails: kirkp@mail.utexas.edu;  
pricet@uchicago.edu

1. Seehausen, O. *et al.* *Nature* **455**, 620–626 (2008).
2. Boughman, J. W. *Nature* **411**, 944–948 (2001).
3. Maan, M. E., Hofker, K. D., van Alphen, J. J. M. & Seehausen, O. *Am. Nat.* **167**, 947–954 (2006).
4. Cummings, M. E. *Evolution* **61**, 530–545 (2007).
5. Genner, M. J. *et al.* *Mol. Biol. Evol.* **24**, 1269–1282 (2007).
6. Seehausen, O. *et al.* *Proc. R. Soc. Lond. B* **270**, 129–137 (2003).
7. Verzijden, M. N. & ten Cate, C. *Biol. Lett.* **3**, 134–136 (2007).
8. ten Cate, C. & Vos, D. R. *Adv. Study Behav.* **28**, 1–31 (1999).
9. Seehausen, O. & Schluter, D. *Proc. R. Soc. Lond. B* **271**, 1345–1353 (2004).

## CLIMATE CHANGE

# When did the icehouse cometh?

Stephen F. Pekar

**The concentration of atmospheric carbon dioxide decreased between 45 million and 25 million years ago, a trend accompanied by glaciation at the poles. Modelling results suggest when and where the ice closed in.**

As atmospheric carbon dioxide is predicted to rise to concentrations not seen in perhaps 25 million years (Myr)<sup>1</sup>, scientists are working to understand the impact on Earth's climate and ice sheets. This requires a shift in perspective: geologists typically use the present as a key to the past, but in this case the past might well be the key to predicting how climate will change in the future.

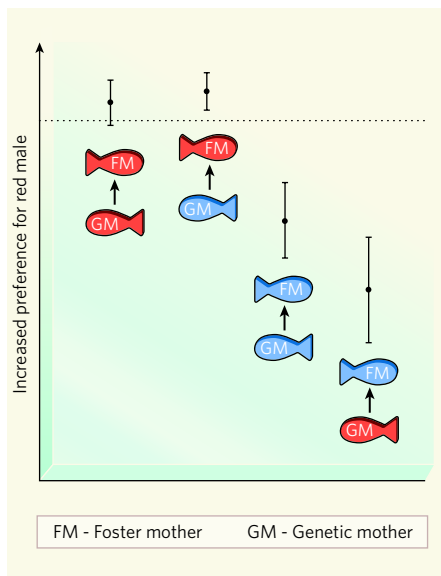
The concentration of CO<sub>2</sub> in the atmosphere is predicted to increase to between 500 and 900 parts per million (p.p.m.) by the end of this century. Geochemical proxies indicate that the last time CO<sub>2</sub> levels were that high was about 45 to 25 Myr ago<sup>1</sup>. This was when Earth changed from a generally ice-free 'greenhouse world' to a more heavily glaciated 'icehouse world'<sup>2,3</sup>, with atmospheric CO<sub>2</sub> gradually decreasing from more than 1,000 p.p.m. to near pre-industrial levels (280 p.p.m.)<sup>1</sup>. So how did the falling atmospheric CO<sub>2</sub> concentrations affect ice-sheet development during this period? On page 652 of this issue, DeConto *et al.*<sup>4</sup> use numerical modelling to constrain the timing of the initiation of glaciation in relation to decreasing levels of CO<sub>2</sub>. Their results not only address a long-standing geological debate, but are also relevant to today's discussion about climate change.

Given the current interest in the effects of CO<sub>2</sub> on climate, it may be surprising to learn that there is a great deal of uncertainty about the extent of ice sheets, and the causal factors in

their development, during the last period when atmospheric CO<sub>2</sub> concentrations reached levels as high as those predicted for the end of this century. Some information can be gleaned by studying the remains of shells from foraminifers — single-celled marine organisms — of that period. The ratio of oxygen isotopes in the shells depends on both the temperature and the isotopic composition of the water in which the foraminifers lived. The isotopic composition, in turn, was controlled by the ice volume at the poles, and by the evaporation–precipitation history of the water when it was near the ocean's surface. By contrast, the ratio of magnesium to calcium in the shells is controlled mainly by the seawater temperature alone. By measuring the two ratios, the isotopic composition of sea water can be calculated and used to constrain the polar ice volume for the period in which the foraminifers were alive.

The data suggest<sup>5,6</sup> that the ice volume was much larger than could be reasonably placed on the Antarctic continent. These high ice-volume estimates, combined with evidence of ice-rafted debris off the coast of Greenland, raise the possibility that glaciation in the Northern Hemisphere might have developed about 40 Myr earlier than was previously thought (that is, up to 44 Myr ago, rather than 3 Myr ago).

DeConto *et al.*<sup>4</sup> cast fresh light on this issue. They developed a model of global climate and of ice-sheet formation that incorporates the decreasing levels of atmospheric CO<sub>2</sub> found



**Figure 2 | The mother who raised you determines with whom you mate.** In choice tests<sup>7</sup> with mouth-brooding cichlids, female fish raised by foster mothers belonging to different species from their genetic mothers preferred males of that different species. The stylized pictures are of males; females are dull and generally similarly coloured. The preference was measured by the difference in the number of approaches per male display when female fish are given a pairwise choice (with standard errors; dashed line indicates random choice). This evidence suggests that learning at a young age contributes to reproductive isolation in cichlids, in addition to other mechanisms such as the action of natural selection on vision described by Seehausen and colleagues<sup>1</sup>. (Modified from ref. 7.)

contributes to speciation above and beyond its effects on mate choice.

But is even this enough? Other findings point to an additional mechanism that complements reproductive isolation via vision. The females of these remarkable fish brood their eggs in their mouths, then guard the young fry after they hatch. In experiments reported last year, Verzijden and ten Cate<sup>7</sup> swapped eggs between the mouths of red morph and blue morph mothers. Females raised from the experimental broods strongly preferred males from their foster morph over those of their own morph (Fig. 2). As females of the two species look very similar, it is unclear whether the offspring preference is based on colour or some other correlated cue such as odour. Regardless of that, learning at a young age (sexual imprinting) apparently contributes to reproductive isolation in these cichlids, as it does in other groups such as birds<sup>8</sup>. The implication is that assortative mating — the tendency of like to mate with like — can arise whenever male characteristics diverge in response to differences in the environment, which might happen even without divergence in the opsin pigments. It remains to be seen if imprinting, vision and perhaps other mechanisms have been sufficient to generate

new species without geographical isolation.

An intriguing observation mentioned by Seehausen *et al.*<sup>1</sup> is that the red- and blue-biased opsin alleles are evolutionarily much older than the species studied here. Red and blue colour morphs are found in other species of cichlid<sup>9</sup>, suggesting that the colour polymorphism may also be ancient. Perhaps one key to the spectacular species radiation of African cichlids is that they inherited from distant ancestors a trove of genetic variation for sensory systems and male signals, possibly contributed during the inferred episode of interbreeding 20,000 years ago. This variation is entrained again and again in speciation events. To systematists, these events represent independent nodes on the evolutionary tree. From the fish's point of view, however, they are perhaps more like an evolutionary play that is re-enacted, night after night, with the same genetic cast. ■

Mark Kirkpatrick is in the Section of Integrative Biology, University of Texas, Austin, Texas 78712, USA. Trevor Price is in the Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA.

e-mails: kirkp@mail.utexas.edu;

pricet@uchicago.edu

1. Seehausen, O. *et al.* *Nature* **455**, 620–626 (2008).
2. Boughman, J. W. *Nature* **411**, 944–948 (2001).
3. Maan, M. E., Hofker, K. D., van Alphen, J. J. M. & Seehausen, O. *Am. Nat.* **167**, 947–954 (2006).
4. Cummings, M. E. *Evolution* **61**, 530–545 (2007).
5. Genner, M. J. *et al.* *Mol. Biol. Evol.* **24**, 1269–1282 (2007).
6. Seehausen, O. *et al.* *Proc. R. Soc. Lond. B* **270**, 129–137 (2003).
7. Verzijden, M. N. & ten Cate, C. *Biol. Lett.* **3**, 134–136 (2007).
8. ten Cate, C. & Vos, D. R. *Adv. Study Behav.* **28**, 1–31 (1999).
9. Seehausen, O. & Schluter, D. *Proc. R. Soc. Lond. B* **271**, 1345–1353 (2004).

## CLIMATE CHANGE

# When did the icehouse cometh?

Stephen F. Pekar

**The concentration of atmospheric carbon dioxide decreased between 45 million and 25 million years ago, a trend accompanied by glaciation at the poles. Modelling results suggest when and where the ice closed in.**

As atmospheric carbon dioxide is predicted to rise to concentrations not seen in perhaps 25 million years (Myr)<sup>1</sup>, scientists are working to understand the impact on Earth's climate and ice sheets. This requires a shift in perspective: geologists typically use the present as a key to the past, but in this case the past might well be the key to predicting how climate will change in the future.

The concentration of CO<sub>2</sub> in the atmosphere is predicted to increase to between 500 and 900 parts per million (p.p.m.) by the end of this century. Geochemical proxies indicate that the last time CO<sub>2</sub> levels were that high was about 45 to 25 Myr ago<sup>1</sup>. This was when Earth changed from a generally ice-free 'greenhouse world' to a more heavily glaciated 'icehouse world'<sup>2,3</sup>, with atmospheric CO<sub>2</sub> gradually decreasing from more than 1,000 p.p.m. to near pre-industrial levels (280 p.p.m.)<sup>1</sup>. So how did the falling atmospheric CO<sub>2</sub> concentrations affect ice-sheet development during this period? On page 652 of this issue, DeConto *et al.*<sup>4</sup> use numerical modelling to constrain the timing of the initiation of glaciation in relation to decreasing levels of CO<sub>2</sub>. Their results not only address a long-standing geological debate, but are also relevant to today's discussion about climate change.

Given the current interest in the effects of CO<sub>2</sub> on climate, it may be surprising to learn that there is a great deal of uncertainty about the extent of ice sheets, and the causal factors in

their development, during the last period when atmospheric CO<sub>2</sub> concentrations reached levels as high as those predicted for the end of this century. Some information can be gleaned by studying the remains of shells from foraminifers — single-celled marine organisms — of that period. The ratio of oxygen isotopes in the shells depends on both the temperature and the isotopic composition of the water in which the foraminifers lived. The isotopic composition, in turn, was controlled by the ice volume at the poles, and by the evaporation–precipitation history of the water when it was near the ocean's surface. By contrast, the ratio of magnesium to calcium in the shells is controlled mainly by the seawater temperature alone. By measuring the two ratios, the isotopic composition of sea water can be calculated and used to constrain the polar ice volume for the period in which the foraminifers were alive.

The data suggest<sup>5,6</sup> that the ice volume was much larger than could be reasonably placed on the Antarctic continent. These high ice-volume estimates, combined with evidence of ice-rafted debris off the coast of Greenland, raise the possibility that glaciation in the Northern Hemisphere might have developed about 40 Myr earlier than was previously thought (that is, up to 44 Myr ago, rather than 3 Myr ago).

DeConto *et al.*<sup>4</sup> cast fresh light on this issue. They developed a model of global climate and of ice-sheet formation that incorporates the decreasing levels of atmospheric CO<sub>2</sub> found



45 to 25 Myr ago, the oxygen-isotope composition of ancient glacial ice, and the expected effects of these parameters on deep-sea records from foraminifers. Their results show that continental-scale Antarctic glaciation would not have developed until CO<sub>2</sub> concentrations reached about 750 p.p.m. — which occurred in the early Oligocene period, 34 to 32 Myr before present (Fig. 1). However, they also predict that the threshold for significant ice-sheet development in the Northern Hemisphere is much lower (280 p.p.m.), and would have occurred about 25 Myr ago.

The authors' results show that, for glaciation to have occurred in the Northern Hemisphere, the drop in CO<sub>2</sub> at the start of the Oligocene must have resulted in CO<sub>2</sub> concentrations far below those estimated by geochemical proxies. They conclude that a unipolar glacial world developed for the first time about 34 Myr ago, coeval with a decrease in water temperature at the sea bottom that was not registered in previously recorded proxy data. These findings are supported by new data from pristinely preserved foraminiferal shells of that period, which show that significant bottom-water cooling must have occurred at the same time that Antarctic ice sheets grew to near modern-day volumes<sup>7</sup>. The new foraminiferal data are also in good agreement with stratigraphic records

of sea-level change from sediments that were deposited on mid- to low-latitude continental margins<sup>8,9</sup>.

DeConto and colleagues<sup>4</sup> also show that, at the CO<sub>2</sub> concentrations that occurred during the middle to late Eocene epoch (45 to 34 Myr ago), small, ephemeral ice sheets could have existed on the highlands of Antarctica — even though CO<sub>2</sub> concentrations were up to six times those of pre-industrial levels. Their conclusions are consistent with ice-volume estimates from stratigraphic records from non-polar continental margins<sup>9,10</sup>. Similarly, the authors demonstrate that small, isolated sheets of glacial ice could have formed in the Northern Hemisphere during the cooler intervals of the Eocene and Oligocene, especially during periods when variations in Earth's orbit produced relatively cold northern summers. This could explain why ice-rafted debris existed off the coast of Greenland during the late-middle Eocene (about 44 Myr ago)<sup>6</sup> without having to invoke the presence of massive continental ice sheets. The transient glacial ice in the Northern Hemisphere might have left a sedimentary record, but would have had insufficient volume to be detectable in oxygen-isotope records.

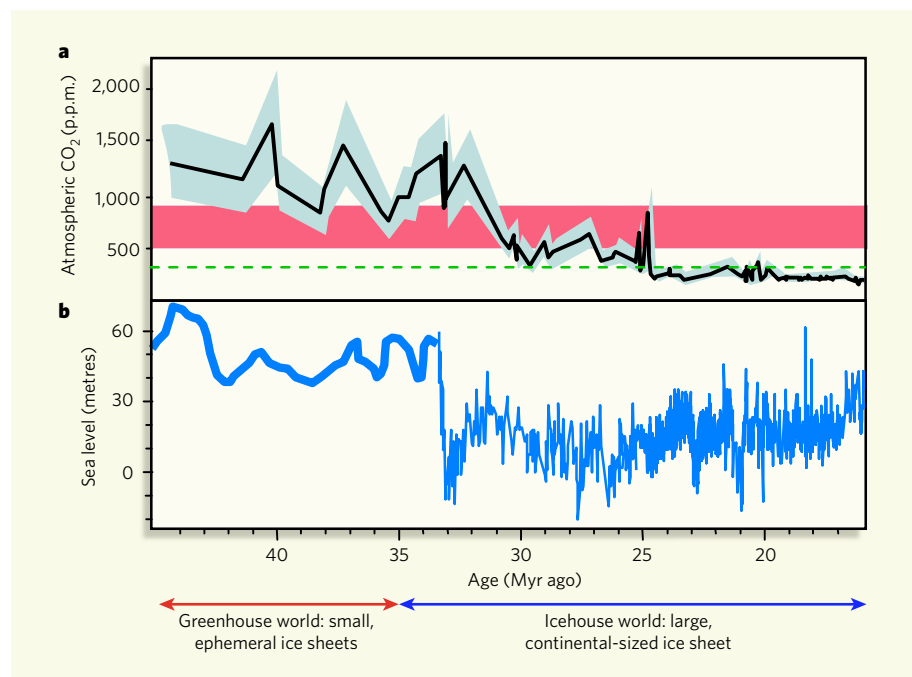
One of DeConto and colleagues' more intriguing conclusions is that, once CO<sub>2</sub> reached near present-day concentrations (about

25 Myr ago), large ice sheets could develop in the Northern Hemisphere during favourable orbits of the Earth. Given that the East Antarctic Ice Sheet is believed to have responded little to changes in climate once it reached near-continental size<sup>11,12</sup>, the large variations in ice-volume indicated in isotopic and stratigraphic records<sup>8</sup> younger than 25 Myr old could therefore be explained in part by episodic ice-sheet growth in the Northern Hemisphere. This suggests that substantial ice growth in the Northern Hemisphere might have started up to 20 Myr earlier than previously believed — up to 25 Myr ago, rather than in the Miocene to the Pleistocene epochs 7 to 3 Myr ago. But there is currently scant evidence to suggest that large amounts of glacial ice existed in the Northern Hemisphere before the late Miocene. In addition, the large changes in ice volume suggested by proxy data for the period in which CO<sub>2</sub> reached near-present-day concentrations (less than 25 Myr ago) do not bode well for long-term sea-level changes in the future, as they suggest that small variations in CO<sub>2</sub> might have large effects on ice volume.

With such a paucity of information about the timing of Earth's ice development (especially in the Northern Hemisphere), there is a clear need for data from high-latitude sites in the Northern Hemisphere and around Antarctica to test DeConto and colleagues' conclusions. Fortunately, help is on its way, as several upcoming projects<sup>13–15</sup> will target strata from the greenhouse–icehouse transition in Antarctica, and are expected to provide additional insight into ice-sheet development in the region during this critical time interval.

Stephen F. Pekar is at the School of Earth and Environmental Sciences, Queens College, 65–30 Kissena Boulevard, Flushing, New York 11367, USA, and at the Lamont Doherty Earth Observatory of Columbia University, New York.  
e-mail: stephen.pekar@qc.cuny.edu

- Pagani, M., Zachos, J. C., Freeman, K. H., Tipler, B. & Bohaty, S. *Science* **309**, 600–603 (2005).
- Miller, K. G., Wright, J. D. & Fairbanks, R. G. *J. Geophys. Res.* **96**, 6829–6848 (1991).
- Lear, C. H., Rosenthal, Y., Coxall, H. K. & Wilson, P. A. *Palaeogeography* doi:10.1029/2004PA001039 (2004).
- DeConto, R. M. *et al. Nature* **455**, 652–656 (2008).
- Coxall, H. K., Wilson, P. A., Pälike, H., Lear, C. H. & Backman, J. *Nature* **433**, 53–57 (2005).
- Tripathi, A. K. *et al. Earth Planet. Sci. Lett.* **265**, 112–122 (2008).
- Lear, C. H., Bailey, T. R., Pearson, P. N., Coxall, H. K. & Rosenthal, Y. *Geology* doi:10.1130/G24584A.1 (2008).
- Pekar, S. F. & Christie-Blick, N. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **260**, 41–49 (2008).
- Miller, K. G. *et al. Science* **310**, 1293–1298 (2005).
- Pekar, S. F., Hucks, A., Fuller, M. & Li, S. *Geol. Soc. Am. Bull.* **117**, 1081–1093 (2005).
- Pollard, D. & DeConto, R. M. *Glob. Planet. Change* **45**, 9–21 (2005).
- Huybrechts, P. *Ann. Glaciol.* **20**, 336–340 (1994).
- [http://qcpages.qc.cuny.edu/offshore\\_new\\_harbor/index.htm](http://qcpages.qc.cuny.edu/offshore_new_harbor/index.htm)
- <http://andriil.org/science/ch>
- [http://iodp.tamu.edu/scienceops/expeditions/wilkes\\_land.html](http://iodp.tamu.edu/scienceops/expeditions/wilkes_land.html)
- Watson, R. T. *et al. Climate Change 2001: Synthesis Report* 184 (IPCC, 2001).



**Figure 1 | Atmospheric carbon dioxide at the start of large-scale glaciation.** About 34 million years (Myr) ago, Earth changed from a greenhouse world (which was generally ice-free) to an icehouse world (which was heavily glaciated). **a**, The concentration of atmospheric CO<sub>2</sub> during this period declined drastically, reaching pre-industrial levels (280 p.p.m., dashed line) about 25 Myr ago. Grey shaded areas represent the uncertainty in the estimates; the red shading shows the range of CO<sub>2</sub> values predicted for the latter part of this century<sup>16</sup>. Estimates of CO<sub>2</sub> concentrations are taken from ref. 1. **b**, Global sea level shows a downward trend. Zero represents sea level when Antarctica is fully glaciated, and increasing values indicate higher sea level (that is, lower ice volume). DeConto and colleagues' model<sup>4</sup> of global climate and ice-sheet formation suggests that, when Earth was a greenhouse world, short-lived glacial formation could occur at a small scale. But when Earth entered its icehouse phase, a large, continental-sized ice sheet formed, with a correlated lowering of the sea level. Apparent sea-level (ASL) estimates from 45 to 34 Myr ago are modified from ref. 9; ASL estimates from 34 to 16 Myr ago are modified from ref. 8.

## REGENERATIVE MEDICINE

## Short cut to cell replacement

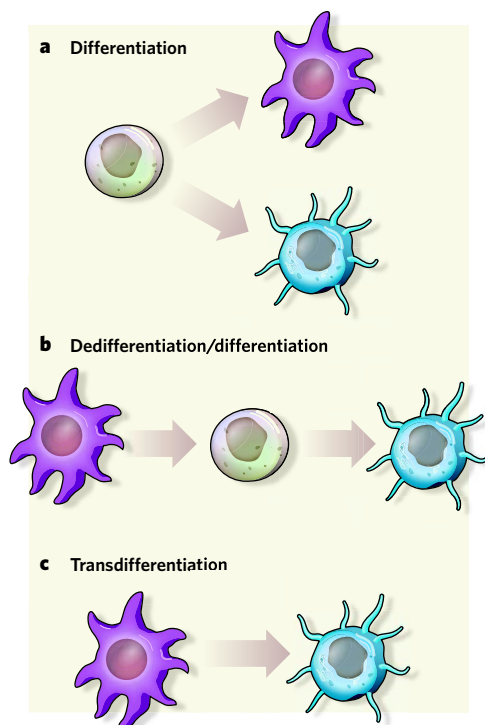
Robert Blelloch

**To make one differentiated cell type from another, a 'stopover' at an undifferentiated state is often required. An alternative method offering an efficient direct route could have implications for disease treatment.**

Regenerative medicine aims to repair diseased or damaged tissues by replacing the affected cells with healthy, functional cells of the same type. The prospects of this discipline have been boosted by the promise of embryonic stem (ES) cells, which are pluripotent — that is, they can differentiate into any cell type — and which can be maintained in culture to 'self-renew' indefinitely. Indeed, recent breakthroughs both in the production of patient-specific ES-like cells and in inducing the differentiation of ES cells into functional adult tissues have provided further hope<sup>1</sup>. Like all promising therapies, however, the use of ES cells has its challenges, among them the difficulties associated with efficiently transplanting and integrating the generated tissue into the physiological framework of the body. On page 627 of this issue, Zhou *et al.*<sup>2</sup> describe an approach whereby differentiated adult cells of one type can be directly and efficiently converted into functional cells of another type within an organism and without the need to first reprogram them into an ES-cell-like state.

As an organism develops, its cells become increasingly specialized, losing their developmental potential (Fig. 1a). This differentiation process involves silencing of gene networks that are no longer needed and activation of other specific networks. These networks are regulated at various levels. First, the RNA and protein composition of a cell drives a specific gene-expression program that is often self-reinforcing. Second, the packaging of DNA — the epigenetic program — affects the access of gene transcription factors to specific genomic regions. Finally, in exceptional cases, the DNA sequence itself can be irreversibly altered. Each of these mechanisms further locks down the developmental potential of the differentiating cell.

For many years, it was presumed that once a cell differentiates, it burns all bridges behind it. But the discovery, first in amphibians and then mammals, that a fully differentiated cell can be manipulated to 'dedifferentiate' (Fig. 1b), and revert to a state resembling that of an early embryonic cell, proved this presumption incorrect. Dedifferentiation can be triggered by either placing the nucleus of a differentiated cell in the cytoplasmic milieu of an egg cell<sup>3</sup> or — as was shown relatively recently<sup>4</sup> — by introducing just four specific transcription factors into the differentiated cell. The latter finding has given a huge boost to the field of regenerative medicine as it indicates that the production of



**Figure 1 | The regenerative-medicine toolbox.** **a**, During development, non-specialized cells with a broad developmental potential differentiate into various highly specialized cells that have limited developmental potential. **b**, Nonetheless, in the lab, these highly specialized cells can be induced to dedifferentiate — that is, revert back to an earlier stem-cell fate with a broad developmental potential. The cells generated in this way can then be triggered to differentiate into another cell type. **c**, Alternatively, in some circumstances, as Zhou *et al.*<sup>2</sup> show, a highly specialized cell can be induced to transdifferentiate into another specialized cell, bypassing the intermediate step of dedifferentiation.

pluripotent cells from many sources, including patients with specific diseases<sup>5,6</sup>, can be relatively straightforward. A crucial goal now is to ensure the safety of such induced pluripotent stem cells and to differentiate them into cells that can be used to repair damaged tissue.

Transdifferentiation — the direct conversion of one differentiated cell type to another (Fig. 1c) — provides an alternative strategy for repairing damaged tissue. In the early 2000s, this approach received a lot of publicity when several reports suggested that transdifferentiation occurs spontaneously across a wide range of tissues<sup>7</sup>. It was soon realized, however, that

what was perceived as spontaneous transdifferentiation was, in fact, the result of cell fusion<sup>7</sup>. Although this realization dampened the general excitement, it did not deter many researchers, who continued to pursue the possibility of inducing transdifferentiation through the introduction of specific 'master regulators' to cells.

In the field of diabetes, for example, efforts were concentrated on generating pancreatic insulin-producing  $\beta$ -islet cells by inducing transdifferentiation of liver cells<sup>8</sup>. In 2003, two groups reported<sup>9,10</sup> that when the gene for either of the transcription factors Pdx1 or NeuroD1 — the latter was used together with the growth factor betacellulin — is directly introduced into the liver of adult mice (using adenoviruses as gene vectors), liver cells transform into long-lived insulin-producing cells. Moreover, these transdifferentiated cells could correct high blood-glucose levels following chemically induced injury of  $\beta$ -islet cells. Nonetheless, the efficiency of this approach was low, and it was unclear which liver-cell type underwent transdifferentiation.

Zhou *et al.*<sup>2</sup> now elegantly marry previous approaches used for inducing transdifferentiation with those that led to the discovery of the four main transcription factors required for dedifferentiation<sup>4</sup>. The outcome is efficient production of  $\beta$ -islet-like cells from a distinct, highly specialized cell type in the pancreas known as exocrine cells.

Using previous data from many labs<sup>11</sup>, the authors identified nine genes that are essential to the embryonic development of  $\beta$ -cells. They used adenoviruses to introduce different combinations of these nine genes into the pancreas of adult mice and found that the introduction of a set of three transcription factors (Ngn3, Pdx1 and Mafa) induces transdifferentiation of an impressive 20% of the manipulated exocrine cells into  $\beta$ -islet-like cells. The authors used a combination of cellular markers and lineage-tracing experiments to prove that the cells that underwent transdifferentiation were, indeed, exocrine cells.

Zhou and colleagues' transdifferentiated  $\beta$ -islet cells resemble these cells' natural counterparts in several important aspects, including the proteins they express; their morphology; and their ability to secrete active insulin, which could diminish high blood-glucose levels following chemically induced pancreatic injury. However, unlike their natural counterparts, these cells do not form, or become incorporated into, islets — islands consisting of  $\beta$ -islet cells, other endocrine cells and blood vessels. Nonetheless, they do form intimate contacts with blood vessels, presumably allowing them to sense blood-glucose levels and release insulin into the circulation accordingly.

These exciting results lead to many intriguing questions. For instance, does transdifferentiation involve epigenetic reprogramming or is the nuclear content of exocrine cells already permissive for the activation of the transcriptional program required for  $\beta$ -islet-cell



formation? And how developmentally distant can the cell of origin and the target cell be for this approach to work? Exocrine cells and  $\beta$ -islet cells share a common precursor. When Zhou *et al.* used the same factors to induce transdifferentiation of more distantly related fibroblast cells or muscle cells into the  $\beta$ -islet-like cells, the approach failed. Could it be that the inclusion of other factors allows transdifferentiation of more distantly related cells?

Considering the many problems associated with the use of viral vectors in gene therapy<sup>12</sup>, finding alternative ways to induce the expression of the three transcription factors in the exocrine cells of an organism would be especially useful. Can the virus-mediated introduction of genes for Ngn3, Pdx1 and Mafa

be substituted with transient introduction of non-DNA elements such as RNAs, recombinant proteins or chemical mimics of these factors, which may be safer? Indeed, the transdifferentiation process that Zhou *et al.* describe occurred rapidly and did not require ongoing expression of the virally introduced genes, suggesting that transient non-DNA-based approaches might succeed. No matter what the answers to these questions are, the authors' findings<sup>2</sup> remind us that the field of regenerative medicine must pursue several strategies to uncover the best therapeutic solutions to degenerative diseases.

Robert Blelloch is at the Institute for Regeneration Medicine, Center for Reproductive Sciences, and Department of Urology, University of California, San Francisco, San Francisco,

California 94143-0525, USA.

e-mail: blellochr@stemcell.ucsf.edu

1. Murry, C. E. & Keller, G. *Cell* **132**, 661–680 (2008).
2. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *Nature* **455**, 627–632 (2008).
3. Hochedlinger, K. & Jaenisch, R. *Nature* **441**, 1061–1067 (2006).
4. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
5. Park, I.-H. *et al.* *Cell* **134**, 877–886 (2008).
6. Dimos, J. T. *et al.* *Science* **321**, 1218–1221 (2008).
7. Rodic, N., Rutenberg, M. S. & Terada, N. *Trends Mol. Med.* **10**, 93–96 (2004).
8. Meivar-Levy, I. & Ferber, S. *Trends Endocrinol. Metab.* **14**, 460–466 (2003).
9. Ber, I. *et al.* *J. Biol. Chem.* **278**, 31950–31957 (2003).
10. Kojima, H. *et al.* *Nature Med.* **9**, 596–603 (2003).
11. Oliver-Krasinski, J. M. & Stoffers, D. A. *Genes Dev.* **22**, 1998–2021 (2008).
12. Hasbrouck, N. C. & High, K. A. *Gene Therapy* **15**, 870–875 (2008).

## AIDS

# Prehistory of HIV-1

Paul M. Sharp and Beatrice H. Hahn

The origin of the current AIDS pandemic has been a subject of great interest and speculation. Viral archaeology sheds light on the geography and timescale of the early diversification of HIV-1 in humans.

Human immunodeficiency virus type 1 (HIV-1) must have been spreading through the human population long before AIDS was first described in 1981, but very few strains from this 'prehistoric' period (pre-1980s) have been characterized. Viral sequences from earlier times can provide insight into the early spread of HIV-1, because the rapid rate of evolution of this virus — up to a million times faster than that of animal DNA — means that substantial amounts of sequence change occur in a matter of decades<sup>1</sup>. On page 661 of this issue, Worobey *et al.*<sup>2</sup> describe the sequences of partial genome fragments of HIV-1 from a lymph-node biopsy collected in 1960 in Léopoldville (now Kinshasa, Democratic Republic of the Congo). They compare these sequences with those of other HIV-1 strains, shedding light on the early evolution and diversification of this virus in Africa.

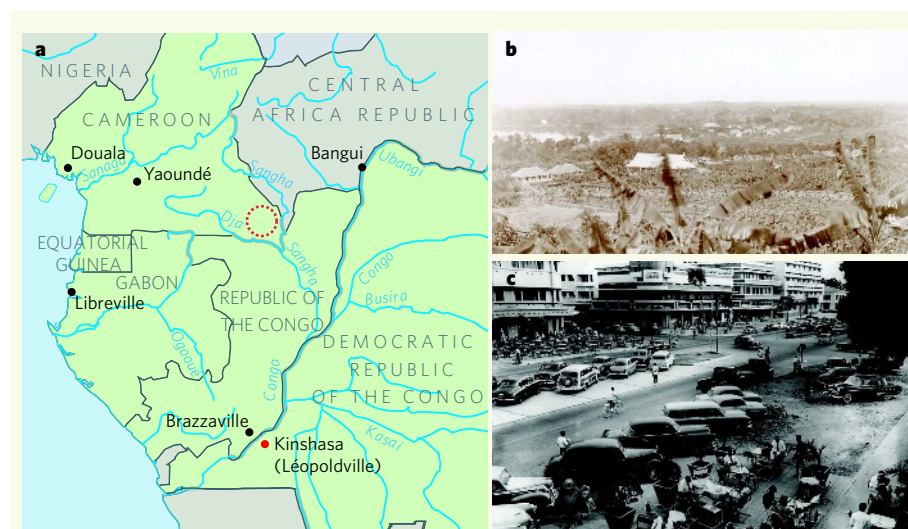
HIV-1 strains are divided into three groups, each of which was independently derived from a simian immunodeficiency virus (SIV) that naturally infects chimpanzees in west-central Africa<sup>3</sup>. Whereas two of these groups are rare, the third, group M, has spread throughout the world and is the cause of more than 95% of HIV infections globally. Group M can be further divided into many subtypes (A–K), which seem to have arisen through founder events. For example, subtype B, which encompasses all the strains originally described in North America and Europe, is very rare in Africa, and reflects such a founder event. Last year, Worobey and colleagues showed<sup>4</sup> that this subtype probably arose from a single strain

that was carried from Africa to Haiti before spreading to the United States and onwards. The newly described<sup>2</sup> 1960 virus (DRC60) falls within, but close to the ancestor of, subtype A.

DRC60 is not the first 'ancient' HIV-1 sample to be characterized: viral sequences from a blood-plasma sample originally obtained in 1959 — also from Léopoldville — were published 10 years ago<sup>5</sup>. The importance of DRC60

is that it is highly divergent from the 1959 sample (ZR59), which was most closely related to the ancestor of subtype D, thus directly demonstrating that, by 50 years ago, group M HIV-1 strains had already undergone substantial diversification.

The ZR59 and DRC60 sequences differ by about 12%, a value similar to distances now seen between the most divergent strains within subtypes. As the positions of ZR59 and DRC60 within the group M phylogeny indicate that the various subtypes already existed 50 years ago, simple extrapolation suggests that these two viral sequences had a common ancestor at least 50 years before that. For a more robust estimate of the date of the common ancestor of HIV-1 group M strains, Worobey and colleagues used state-of-the-art statistical analyses, allowing a variety of models for the growth of the HIV-1 pandemic and variable rates of evolution. The different analyses gave broadly similar



**Figure 1 | Origin of pandemic HIV-1.** **a**, Map of west-central Africa showing major rivers, and cities with explosive population growth in the twentieth century. Chimpanzees carrying the SIV strains most closely related to the viruses of HIV-1 group M, such as that described by Worobey *et al.*<sup>2</sup>, have been found in southeast Cameroon (red ring). **b**, Léopoldville in 1896 (view from Mount Léopold) and **c**, around 1955 (the commercial centre). (Photos F. L. Michel (**b**) and C. Lamote (**c**), collection of the Royal Museum for Central Africa, Tervuren, Belgium.)

formation? And how developmentally distant can the cell of origin and the target cell be for this approach to work? Exocrine cells and  $\beta$ -islet cells share a common precursor. When Zhou *et al.* used the same factors to induce transdifferentiation of more distantly related fibroblast cells or muscle cells into the  $\beta$ -islet-like cells, the approach failed. Could it be that the inclusion of other factors allows transdifferentiation of more distantly related cells?

Considering the many problems associated with the use of viral vectors in gene therapy<sup>12</sup>, finding alternative ways to induce the expression of the three transcription factors in the exocrine cells of an organism would be especially useful. Can the virus-mediated introduction of genes for Ngn3, Pdx1 and Mafa

be substituted with transient introduction of non-DNA elements such as RNAs, recombinant proteins or chemical mimics of these factors, which may be safer? Indeed, the transdifferentiation process that Zhou *et al.* describe occurred rapidly and did not require ongoing expression of the virally introduced genes, suggesting that transient non-DNA-based approaches might succeed. No matter what the answers to these questions are, the authors' findings<sup>2</sup> remind us that the field of regenerative medicine must pursue several strategies to uncover the best therapeutic solutions to degenerative diseases.

Robert Blelloch is at the Institute for Regeneration Medicine, Center for Reproductive Sciences, and Department of Urology, University of California, San Francisco, San Francisco,

California 94143-0525, USA.

e-mail: blellochr@stemcell.ucsf.edu

1. Murry, C. E. & Keller, G. *Cell* **132**, 661–680 (2008).
2. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *Nature* **455**, 627–632 (2008).
3. Hochedlinger, K. & Jaenisch, R. *Nature* **441**, 1061–1067 (2006).
4. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
5. Park, I.-H. *et al.* *Cell* **134**, 877–886 (2008).
6. Dimos, J. T. *et al.* *Science* **321**, 1218–1221 (2008).
7. Rodic, N., Rutenberg, M. S. & Terada, N. *Trends Mol. Med.* **10**, 93–96 (2004).
8. Meivar-Levy, I. & Ferber, S. *Trends Endocrinol. Metab.* **14**, 460–466 (2003).
9. Ber, I. *et al.* *J. Biol. Chem.* **278**, 31950–31957 (2003).
10. Kojima, H. *et al.* *Nature Med.* **9**, 596–603 (2003).
11. Oliver-Krasinski, J. M. & Stoffers, D. A. *Genes Dev.* **22**, 1998–2021 (2008).
12. Hasbrouck, N. C. & High, K. A. *Gene Therapy* **15**, 870–875 (2008).

## AIDS

# Prehistory of HIV-1

Paul M. Sharp and Beatrice H. Hahn

The origin of the current AIDS pandemic has been a subject of great interest and speculation. Viral archaeology sheds light on the geography and timescale of the early diversification of HIV-1 in humans.

Human immunodeficiency virus type 1 (HIV-1) must have been spreading through the human population long before AIDS was first described in 1981, but very few strains from this 'prehistoric' period (pre-1980s) have been characterized. Viral sequences from earlier times can provide insight into the early spread of HIV-1, because the rapid rate of evolution of this virus — up to a million times faster than that of animal DNA — means that substantial amounts of sequence change occur in a matter of decades<sup>1</sup>. On page 661 of this issue, Worobey *et al.*<sup>2</sup> describe the sequences of partial genome fragments of HIV-1 from a lymph-node biopsy collected in 1960 in Léopoldville (now Kinshasa, Democratic Republic of the Congo). They compare these sequences with those of other HIV-1 strains, shedding light on the early evolution and diversification of this virus in Africa.

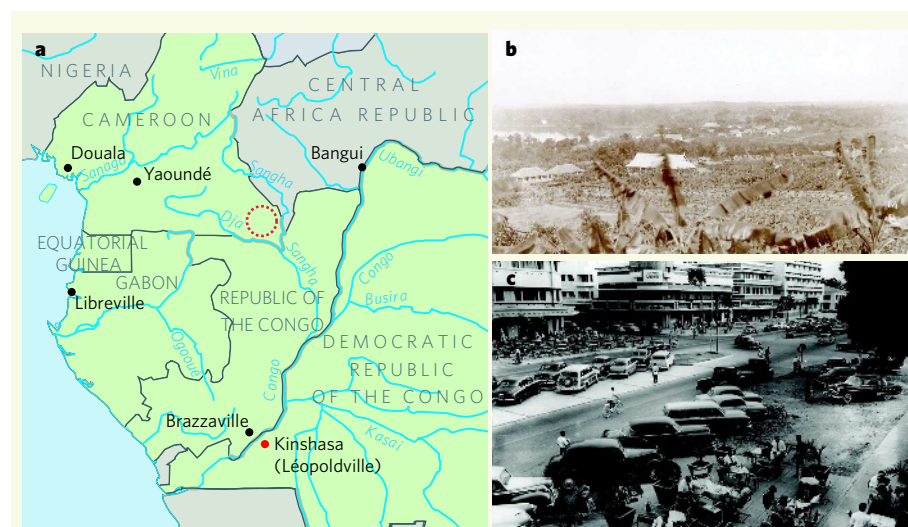
HIV-1 strains are divided into three groups, each of which was independently derived from a simian immunodeficiency virus (SIV) that naturally infects chimpanzees in west-central Africa<sup>3</sup>. Whereas two of these groups are rare, the third, group M, has spread throughout the world and is the cause of more than 95% of HIV infections globally. Group M can be further divided into many subtypes (A–K), which seem to have arisen through founder events. For example, subtype B, which encompasses all the strains originally described in North America and Europe, is very rare in Africa, and reflects such a founder event. Last year, Worobey and colleagues showed<sup>4</sup> that this subtype probably arose from a single strain

that was carried from Africa to Haiti before spreading to the United States and onwards. The newly described<sup>2</sup> 1960 virus (DRC60) falls within, but close to the ancestor of, subtype A.

DRC60 is not the first 'ancient' HIV-1 sample to be characterized: viral sequences from a blood-plasma sample originally obtained in 1959 — also from Léopoldville — were published 10 years ago<sup>5</sup>. The importance of DRC60

is that it is highly divergent from the 1959 sample (ZR59), which was most closely related to the ancestor of subtype D, thus directly demonstrating that, by 50 years ago, group M HIV-1 strains had already undergone substantial diversification.

The ZR59 and DRC60 sequences differ by about 12%, a value similar to distances now seen between the most divergent strains within subtypes. As the positions of ZR59 and DRC60 within the group M phylogeny indicate that the various subtypes already existed 50 years ago, simple extrapolation suggests that these two viral sequences had a common ancestor at least 50 years before that. For a more robust estimate of the date of the common ancestor of HIV-1 group M strains, Worobey and colleagues used state-of-the-art statistical analyses, allowing a variety of models for the growth of the HIV-1 pandemic and variable rates of evolution. The different analyses gave broadly similar



**Figure 1 | Origin of pandemic HIV-1.** **a**, Map of west-central Africa showing major rivers, and cities with explosive population growth in the twentieth century. Chimpanzees carrying the SIV strains most closely related to the viruses of HIV-1 group M, such as that described by Worobey *et al.*<sup>2</sup>, have been found in southeast Cameroon (red ring). **b**, Léopoldville in 1896 (view from Mount Léopold) and **c**, around 1955 (the commercial centre). (Photos F. L. Michel (**b**) and C. Lamote (**c**), collection of the Royal Museum for Central Africa, Tervuren, Belgium.)



estimates for the date of that common ancestor, between 1902 and 1921, with 95% confidence intervals ranging no later than 1933. These dates are a little earlier than, but do not differ significantly from, a previous estimate<sup>1</sup> of 1931 from an analysis that did not include the 50-year-old viruses.

The interpretation that HIV-1 was spreading among humans for 60–80 years before AIDS was first recognized should not be surprising. If the epidemic grew roughly exponentially from only one or a few infected individuals around 1910 to the more than 55 million estimated to have been infected by 2007, there were probably only a few thousand HIV-infected individuals by 1960, all in central Africa. Given the diverse array of symptoms characteristic of AIDS, and the often-long asymptomatic period following infection, it is easy to imagine how the nascent epidemic went unrecognized. Conversely, such a low prevalence at that time implies that the Congolese co-authors of the paper<sup>2</sup> were very lucky to come across this infected sample, even if most infections were concentrated in the area of Léopoldville. But can we trust these sequences?

In work on ancient DNA, contamination is especially problematic, and the work should, if possible, be replicated in other laboratories. For DRC60, independent analyses were performed at the University of Arizona and Northwestern University, Illinois. The sequences obtained were similar, but not identical, exactly as expected when samples come from the diverse set of related viral sequences that — because of the virus's rapid rate of evolution — arise within an infected individual<sup>6</sup>. Furthermore, the distance along the evolutionary tree from the group M ancestor to the ZR59 or DRC60 sequences is much shorter than those between the ancestor and modern strains, consistent with the earlier dates of isolation of ZR59 and DRC60, and confirming that these viruses are indeed old.

Although the ZR59 and DRC60 sequences can show only that two subtypes were present in Léopoldville around 1960, in more recent times the greatest diversity of group M subtypes — as well as many divergent strains that have not been classified — has been found in Kinshasa<sup>7</sup>. So it seems likely that all of the early diversification of HIV-1 group M viruses occurred in the Léopoldville area. Yet the SIV strains most closely related to HIV-1 group M have been found infecting chimpanzees in the southeast corner of Cameroon<sup>3</sup>, some 700 kilometres away (Fig. 1a). The simplest explanation for how SIV jumped to humans would be through exposure of humans to the blood of chimpanzees butchered locally for bushmeat. So why did the pandemic start in Léopoldville? And, as there must have been many opportunities for such transmission over past millennia, why did the AIDS pandemic not occur until the twentieth century?

The answer may be that, for an AIDS epidemic to get kick-started, HIV-1 needs to

be seeded in a large population centre. But cities of significant size did not exist in central Africa before 1900. Worobey and colleagues<sup>2</sup> reproduce demographic data showing the rapid growth of cities in west-central Africa during the twentieth century. Léopoldville was not only the largest of these cities, but also a likely destination for a virus escaping from southeast Cameroon. In the early 1900s, the main routes of transportation out of that remote forest region were rivers; those surrounding this area flow south, ultimately draining into the Congo River, and leading to Léopoldville (Fig. 1).

The date estimates of Worobey *et al.* are for an ancestral virus, present in the first individual to give rise to separate transmission chains that still exist today. We may never know how many individuals were infected in the previous transmission chain, the one that led from the person initially infected with SIV

to the progenitor of the current pandemic in humans. This exception aside, we can now paint a remarkably detailed picture of the time and place of origin of HIV-1 group M viruses and their early diversification, and thus of the prehistory of the AIDS pandemic. ■

Paul M. Sharp is at the Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. Beatrice H. Hahn is in the Departments of Medicine and Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA.  
e-mails: paul.sharp@ed.ac.uk; bhahn@uab.edu

1. Korber, B. *et al.* *Science* **288**, 1789–1796 (2000).
2. Worobey, M. *et al.* *Nature* **455**, 661–664 (2008).
3. Keele, B. F. *et al.* *Science* **313**, 523–526 (2006).
4. Gilbert, M. T. P. *et al.* *Proc. Natl Acad. Sci. USA* **104**, 18566–18570 (2007).
5. Zhu, T. *et al.* *Nature* **391**, 594–597 (1998).
6. Meyerhans, A. *et al.* *Cell* **58**, 901–910 (1989).
7. Vidal, N. *et al.* *J. Virol.* **74**, 10498–10507 (2000).

## APPLIED PHYSICS

# Virtues of diamond defects

Michael Romalis

**A general method for detecting nuclear magnetic resonance signals from a single molecule has so far been elusive. Magnetic sensors that exploit crystal imperfections in diamond might make such a method a reality.**

The phenomenon of nuclear magnetic resonance (NMR), which results from the interaction of the spin of an atomic nucleus with an external magnetic field, has successfully been exploited in such disparate techniques as the structural analysis of molecules (NMR spectroscopy) and structural and functional analysis of the human body (NMR imaging), thus spanning length-scales from ångströms to metres. But these techniques have remained mostly bulk methods, in that they usually require more than a billion spins. Writing in this issue, Maze *et al.*<sup>1</sup> (page 644) and Balasubramanian *et al.*<sup>2</sup> (page 648) describe a new approach to NMR detection that exploits a single spin associated with a crystal imperfection — a 'nitrogen-vacancy centre' — in diamond to achieve unprecedented magnetic-field sensitivity on the nanometre scale. Crucially, the approach works at room temperature, a key requirement for biological applications.

The ideal imaging technique would be passive, applicable to all materials, and have spatial resolution on the atomic scale. In fact, nature provides us with the means for achieving such imaging. Electrons and many nuclei have a finite magnetic moment associated with their spins, and thus create weak magnetic fields around themselves. All one needs, then, is a magnetic sensor with high enough sensitivity to measure these fields.

Several magnetometry techniques have been considered for this purpose, but they

mostly fall short of the required sensitivity. The sensitivity of existing magnetic-field sensors improves with their characteristic length-scale ( $r$ ) approximately to the power  $3/2$  ( $r^{3/2}$ ). But the magnetic field generated by the magnetic moment of a single electron or proton drops with the third power of the distance from the spin ( $r^{-3}$ ). Thus, the best hope for detecting magnetic fields from single spins comes from sensors on the nanometre scale.

A milestone in this direction was achieved a few years ago when a cantilever with a magnetic tip was used to detect the magnetic field created by the spin of a single electron<sup>3</sup>. Further improvements to this approach, known as magnetic resonance force microscopy (MRFM), might allow detection of the magnetic field produced by a single nuclear spin, which is a thousand times smaller than that produced by an electron's spin. But MRFM remains a challenging method because it requires a cryogenic environment and prolonged data averaging.

Optical methods involving scattering of light have long been used to study single particles. This is because it is relatively straightforward to detect individual photons. In some atomic systems, the scattering of photons can be made to depend on the direction of the particle's spin, allowing individual spins to be detected. Such methods have been used to detect electron magnetic resonance signals from a single optically active molecule<sup>4,5</sup>, but they cannot be used

estimates for the date of that common ancestor, between 1902 and 1921, with 95% confidence intervals ranging no later than 1933. These dates are a little earlier than, but do not differ significantly from, a previous estimate<sup>1</sup> of 1931 from an analysis that did not include the 50-year-old viruses.

The interpretation that HIV-1 was spreading among humans for 60–80 years before AIDS was first recognized should not be surprising. If the epidemic grew roughly exponentially from only one or a few infected individuals around 1910 to the more than 55 million estimated to have been infected by 2007, there were probably only a few thousand HIV-infected individuals by 1960, all in central Africa. Given the diverse array of symptoms characteristic of AIDS, and the often-long asymptomatic period following infection, it is easy to imagine how the nascent epidemic went unrecognized. Conversely, such a low prevalence at that time implies that the Congolese co-authors of the paper<sup>2</sup> were very lucky to come across this infected sample, even if most infections were concentrated in the area of Léopoldville. But can we trust these sequences?

In work on ancient DNA, contamination is especially problematic, and the work should, if possible, be replicated in other laboratories. For DRC60, independent analyses were performed at the University of Arizona and Northwestern University, Illinois. The sequences obtained were similar, but not identical, exactly as expected when samples come from the diverse set of related viral sequences that — because of the virus's rapid rate of evolution — arise within an infected individual<sup>6</sup>. Furthermore, the distance along the evolutionary tree from the group M ancestor to the ZR59 or DRC60 sequences is much shorter than those between the ancestor and modern strains, consistent with the earlier dates of isolation of ZR59 and DRC60, and confirming that these viruses are indeed old.

Although the ZR59 and DRC60 sequences can show only that two subtypes were present in Léopoldville around 1960, in more recent times the greatest diversity of group M subtypes — as well as many divergent strains that have not been classified — has been found in Kinshasa<sup>7</sup>. So it seems likely that all of the early diversification of HIV-1 group M viruses occurred in the Léopoldville area. Yet the SIV strains most closely related to HIV-1 group M have been found infecting chimpanzees in the southeast corner of Cameroon<sup>3</sup>, some 700 kilometres away (Fig. 1a). The simplest explanation for how SIV jumped to humans would be through exposure of humans to the blood of chimpanzees butchered locally for bushmeat. So why did the pandemic start in Léopoldville? And, as there must have been many opportunities for such transmission over past millennia, why did the AIDS pandemic not occur until the twentieth century?

The answer may be that, for an AIDS epidemic to get kick-started, HIV-1 needs to

be seeded in a large population centre. But cities of significant size did not exist in central Africa before 1900. Worobey and colleagues<sup>2</sup> reproduce demographic data showing the rapid growth of cities in west-central Africa during the twentieth century. Léopoldville was not only the largest of these cities, but also a likely destination for a virus escaping from southeast Cameroon. In the early 1900s, the main routes of transportation out of that remote forest region were rivers; those surrounding this area flow south, ultimately draining into the Congo River, and leading to Léopoldville (Fig. 1).

The date estimates of Worobey *et al.* are for an ancestral virus, present in the first individual to give rise to separate transmission chains that still exist today. We may never know how many individuals were infected in the previous transmission chain, the one that led from the person initially infected with SIV

to the progenitor of the current pandemic in humans. This exception aside, we can now paint a remarkably detailed picture of the time and place of origin of HIV-1 group M viruses and their early diversification, and thus of the prehistory of the AIDS pandemic. ■

Paul M. Sharp is at the Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. Beatrice H. Hahn is in the Departments of Medicine and Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA.  
e-mails: paul.sharp@ed.ac.uk; bhahn@uab.edu

1. Korber, B. *et al.* *Science* **288**, 1789–1796 (2000).
2. Worobey, M. *et al.* *Nature* **455**, 661–664 (2008).
3. Keele, B. F. *et al.* *Science* **313**, 523–526 (2006).
4. Gilbert, M. T. P. *et al.* *Proc. Natl Acad. Sci. USA* **104**, 18566–18570 (2007).
5. Zhu, T. *et al.* *Nature* **391**, 594–597 (1998).
6. Meyerhans, A. *et al.* *Cell* **58**, 901–910 (1989).
7. Vidal, N. *et al.* *J. Virol.* **74**, 10498–10507 (2000).

## APPLIED PHYSICS

# Virtues of diamond defects

Michael Romalis

**A general method for detecting nuclear magnetic resonance signals from a single molecule has so far been elusive. Magnetic sensors that exploit crystal imperfections in diamond might make such a method a reality.**

The phenomenon of nuclear magnetic resonance (NMR), which results from the interaction of the spin of an atomic nucleus with an external magnetic field, has successfully been exploited in such disparate techniques as the structural analysis of molecules (NMR spectroscopy) and structural and functional analysis of the human body (NMR imaging), thus spanning length-scales from ångströms to metres. But these techniques have remained mostly bulk methods, in that they usually require more than a billion spins. Writing in this issue, Maze *et al.*<sup>1</sup> (page 644) and Balasubramanian *et al.*<sup>2</sup> (page 648) describe a new approach to NMR detection that exploits a single spin associated with a crystal imperfection — a 'nitrogen-vacancy centre' — in diamond to achieve unprecedented magnetic-field sensitivity on the nanometre scale. Crucially, the approach works at room temperature, a key requirement for biological applications.

The ideal imaging technique would be passive, applicable to all materials, and have spatial resolution on the atomic scale. In fact, nature provides us with the means for achieving such imaging. Electrons and many nuclei have a finite magnetic moment associated with their spins, and thus create weak magnetic fields around themselves. All one needs, then, is a magnetic sensor with high enough sensitivity to measure these fields.

Several magnetometry techniques have been considered for this purpose, but they

mostly fall short of the required sensitivity. The sensitivity of existing magnetic-field sensors improves with their characteristic length-scale ( $r$ ) approximately to the power  $3/2$  ( $r^{3/2}$ ). But the magnetic field generated by the magnetic moment of a single electron or proton drops with the third power of the distance from the spin ( $r^{-3}$ ). Thus, the best hope for detecting magnetic fields from single spins comes from sensors on the nanometre scale.

A milestone in this direction was achieved a few years ago when a cantilever with a magnetic tip was used to detect the magnetic field created by the spin of a single electron<sup>3</sup>. Further improvements to this approach, known as magnetic resonance force microscopy (MRFM), might allow detection of the magnetic field produced by a single nuclear spin, which is a thousand times smaller than that produced by an electron's spin. But MRFM remains a challenging method because it requires a cryogenic environment and prolonged data averaging.

Optical methods involving scattering of light have long been used to study single particles. This is because it is relatively straightforward to detect individual photons. In some atomic systems, the scattering of photons can be made to depend on the direction of the particle's spin, allowing individual spins to be detected. Such methods have been used to detect electron magnetic resonance signals from a single optically active molecule<sup>4,5</sup>, but they cannot be used



in most samples because they lack convenient optical transitions.

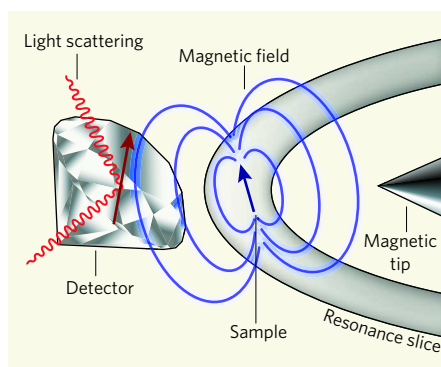
To develop a more general method, it is possible to use one spin with optical readout as a detector of the magnetic fields created by the spins in the sample under analysis (Fig. 1). To achieve sufficient sensitivity, the interaction between spins must persist for a relatively long time. This requires both the detector and the sample spins to have long spin-coherence times, so that their direction is only occasionally perturbed by quantum fluctuations. There are many techniques for achieving long spin-coherence times in atomic systems, even for a single spin that is held, for example, in a laser trap. However, it is generally impossible to bring a sample within nanometres of such spin without disrupting the trapping mechanism.

This is where the properties of nitrogen-vacancy centres in diamond become useful. Recent work<sup>6</sup> has shown that a spin-coherence time of the order of a millisecond can be achieved in such a solid-state system. Even in diamond nanocrystals that are tens of nanometres across, the spin-coherence time is not dramatically reduced. Furthermore, initialization and detection of the spin's direction can be achieved at room temperature, so a diamond magnetometer can be placed within tens of nanometres of a biologically active sample.

Such long spin-coherence times can only be achieved by periodically flipping the direction of the spin, a technique known as spin echo, which averages out external fluctuations. With this technique, which was first used in nitrogen-vacancy centres for quantum-computing applications<sup>7</sup>, Maze *et al.*<sup>1</sup> describe a magnetometer with a single nitrogen-vacancy centre in both a bulk diamond and a nanocrystal. The magnetometer is sensitive to magnetic fields that oscillate at the frequency of the spin-echo repetition rate, typically in the kilohertz range. After collecting light from the nitrogen-vacancy centre for 100 seconds, Maze and colleagues obtained a magnetic-field sensitivity as low as 3 nanotesla, equal to the magnetic field about 100 nanometres from a single electron, or 10 nanometres from a single proton. This is considerably better than has been achieved with other techniques, such as MRFM, on the nanometre scale.

Balasubramanian *et al.*<sup>2</sup> operate a nitrogen-vacancy diamond magnetometer placed next to a magnetic tip, creating a strong magnetic-field gradient. Because the magnetometer relies on a single, well-localized spin, its magnetic resonance is not broadened by the steep spatial variation of the magnetic field. Such a high gradient could allow magnetic-resonance imaging with sub-nanometre resolution. As a first step in this direction, Balasubramanian and colleagues locate the position of the nitrogen-vacancy centre itself with a spatial resolution of 5 nanometres by measuring its resonance frequency.

A combination of the techniques developed by Maze *et al.*<sup>1</sup> and Balasubramanian *et al.*<sup>2</sup> could lead to the detection and imaging of



**Figure 1 | Optical spin detection using a diamond defect in a nanocrystal.** The atomic spin (red arrow) in a crystal defect — a nitrogen-vacancy centre — in a diamond nanoparticle can be manipulated with light. It responds to the magnetic field generated by the spins in a sample, thus changing the light scattering rate. A strong magnetic-field gradient produced by a magnetic tip selects spins with magnetic resonance in a narrow slice of the sample. Maze *et al.*<sup>1</sup> and Balasubramanian *et al.*<sup>2</sup> have implemented parts of this scheme individually. Integration of these techniques could lead to detection of nuclear magnetic resonance from a single spin.

individual nuclear spins, which would potentially allow the determination of the structure of a single molecule. Crucially, both experiments were performed at room temperature, so biological applications such as the determination of protein structures or detailed imaging of the internal structure of a living cell seem feasible.

Manipulation of spin in diamond is a

fast-developing research area, and many avenues remain to be explored. For example, longer spin-coherence times could be achieved using artificial diamond crystals that have a reduced abundance of the carbon-13 isotope, which creates magnetic fields that perturb the sensor. Another point to note is that the NMR signal on very small length-scales is effectively increased because of quantum fluctuations. As the number of spins ( $N$ ) decreases, their quantum fluctuations, which scale as  $N^{1/2}$ , become relatively larger<sup>8</sup>. By resolving individual spins, one can obtain NMR signals that correspond to nearly complete spin polarization even in a low magnetic field, alleviating the current need for strong superconducting magnets in NMR detection, which can polarize only 1 spin in  $10^4$ . Although integration of nitrogen-vacancy spin-echo techniques with the conditions necessary for NMR remains to be demonstrated, diamond magnetometers seem to provide a promising route towards single-spin NMR detection. ■

Michael Romalis is in the Department of Physics, Princeton University, Princeton, New Jersey 08540, USA.

e-mail: romalis@princeton.edu

1. Maze, J. R. *et al.* *Nature* **455**, 644–647 (2008).
2. Balasubramanian, G. *et al.* *Nature* **455**, 648–651 (2008).
3. Rugar, D., Bundaklan, R., Mamin, H. J. & Chui, B. W. *Nature* **430**, 329–332 (2004).
4. Köhler, J. *et al.* *Nature* **363**, 242–244 (1993).
5. Wrachtrup, J. *et al.* *Nature* **363**, 244–245 (1993).
6. Praver, S. & Greentree, A. D. *Science* **320**, 1601–1602 (2008).
7. Childress, L. *et al.* *Science* **314**, 281–285 (2006).
8. Mamin, H. J., Poggio, M., Degen, C. L. & Rugar, D. *Nature Nanotech.* **2**, 301–306 (2007).

## NEUROSCIENCE

# Fragile dopamine

David Weinshenker and Stephen T. Warren

**Dopamine dysfunction, which is implicated in Parkinson's disease and drug addiction, seems an unlikely culprit in fragile X syndrome. A surprising set of findings means a rethink is required.**

Fragile X syndrome is the commonest inherited form of mental retardation, with the patients often also having autism and attention-deficit hyperactivity disorder<sup>1</sup>. It is usually caused by the absence of the protein FMRP, which is encoded by *FMR1*, a gene on the X chromosome. Although FMRP function is not well understood, most studies concur that it is a selective RNA-binding protein that modulates the translation of its target messenger RNAs<sup>2</sup>. But this deceptively simple description of FMRP function omits any role for the neurotransmitter dopamine, despite the fact that some of the clinical and behavioural features of fragile X syndrome are reminiscent of dysfunction in dopamine-secreting neurons. Writing in *Neuron*, Wang *et al.*<sup>3</sup>

elucidate the role of FMRP in modulating dopamine signalling.

Classically, dopamine has been considered an essential mediator of behaviours such as reward-seeking and coordinated movement. Because these brain functions are not always impaired in fragile X syndrome, there has been little reason to link altered dopamine signalling to the aetiology or the manifestation of this disorder. Nevertheless, recent work<sup>4,5</sup> has expanded and refined dopamine's job description in the brain to encompass more cognitively relevant functions, such as involvement in reward-prediction error, motivation, ability to focus on pertinent environmental stimuli and goal-directed behaviours.

To explore a possible connection between

in most samples because they lack convenient optical transitions.

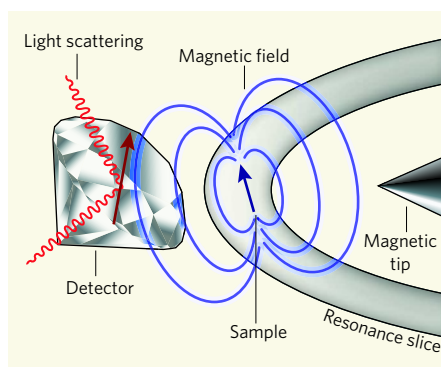
To develop a more general method, it is possible to use one spin with optical readout as a detector of the magnetic fields created by the spins in the sample under analysis (Fig. 1). To achieve sufficient sensitivity, the interaction between spins must persist for a relatively long time. This requires both the detector and the sample spins to have long spin-coherence times, so that their direction is only occasionally perturbed by quantum fluctuations. There are many techniques for achieving long spin-coherence times in atomic systems, even for a single spin that is held, for example, in a laser trap. However, it is generally impossible to bring a sample within nanometres of such spin without disrupting the trapping mechanism.

This is where the properties of nitrogen-vacancy centres in diamond become useful. Recent work<sup>6</sup> has shown that a spin-coherence time of the order of a millisecond can be achieved in such a solid-state system. Even in diamond nanocrystals that are tens of nanometres across, the spin-coherence time is not dramatically reduced. Furthermore, initialization and detection of the spin's direction can be achieved at room temperature, so a diamond magnetometer can be placed within tens of nanometres of a biologically active sample.

Such long spin-coherence times can only be achieved by periodically flipping the direction of the spin, a technique known as spin echo, which averages out external fluctuations. With this technique, which was first used in nitrogen-vacancy centres for quantum-computing applications<sup>7</sup>, Maze *et al.*<sup>1</sup> describe a magnetometer with a single nitrogen-vacancy centre in both a bulk diamond and a nanocrystal. The magnetometer is sensitive to magnetic fields that oscillate at the frequency of the spin-echo repetition rate, typically in the kilohertz range. After collecting light from the nitrogen-vacancy centre for 100 seconds, Maze and colleagues obtained a magnetic-field sensitivity as low as 3 nanotesla, equal to the magnetic field about 100 nanometres from a single electron, or 10 nanometres from a single proton. This is considerably better than has been achieved with other techniques, such as MRFM, on the nanometre scale.

Balasubramanian *et al.*<sup>2</sup> operate a nitrogen-vacancy diamond magnetometer placed next to a magnetic tip, creating a strong magnetic-field gradient. Because the magnetometer relies on a single, well-localized spin, its magnetic resonance is not broadened by the steep spatial variation of the magnetic field. Such a high gradient could allow magnetic-resonance imaging with sub-nanometre resolution. As a first step in this direction, Balasubramanian and colleagues locate the position of the nitrogen-vacancy centre itself with a spatial resolution of 5 nanometres by measuring its resonance frequency.

A combination of the techniques developed by Maze *et al.*<sup>1</sup> and Balasubramanian *et al.*<sup>2</sup> could lead to the detection and imaging of



**Figure 1 | Optical spin detection using a diamond defect in a nanocrystal.** The atomic spin (red arrow) in a crystal defect — a nitrogen-vacancy centre — in a diamond nanoparticle can be manipulated with light. It responds to the magnetic field generated by the spins in a sample, thus changing the light scattering rate. A strong magnetic-field gradient produced by a magnetic tip selects spins with magnetic resonance in a narrow slice of the sample. Maze *et al.*<sup>1</sup> and Balasubramanian *et al.*<sup>2</sup> have implemented parts of this scheme individually. Integration of these techniques could lead to detection of nuclear magnetic resonance from a single spin.

individual nuclear spins, which would potentially allow the determination of the structure of a single molecule. Crucially, both experiments were performed at room temperature, so biological applications such as the determination of protein structures or detailed imaging of the internal structure of a living cell seem feasible.

Manipulation of spin in diamond is a

fast-developing research area, and many avenues remain to be explored. For example, longer spin-coherence times could be achieved using artificial diamond crystals that have a reduced abundance of the carbon-13 isotope, which creates magnetic fields that perturb the sensor. Another point to note is that the NMR signal on very small length-scales is effectively increased because of quantum fluctuations. As the number of spins ( $N$ ) decreases, their quantum fluctuations, which scale as  $N^{1/2}$ , become relatively larger<sup>8</sup>. By resolving individual spins, one can obtain NMR signals that correspond to nearly complete spin polarization even in a low magnetic field, alleviating the current need for strong superconducting magnets in NMR detection, which can polarize only 1 spin in  $10^4$ . Although integration of nitrogen-vacancy spin-echo techniques with the conditions necessary for NMR remains to be demonstrated, diamond magnetometers seem to provide a promising route towards single-spin NMR detection. ■

Michael Romalis is in the Department of Physics, Princeton University, Princeton, New Jersey 08540, USA.

e-mail: romalis@princeton.edu

1. Maze, J. R. *et al.* *Nature* **455**, 644–647 (2008).
2. Balasubramanian, G. *et al.* *Nature* **455**, 648–651 (2008).
3. Rugar, D., Bundaklan, R., Mamin, H. J. & Chui, B. W. *Nature* **430**, 329–332 (2004).
4. Köhler, J. *et al.* *Nature* **363**, 242–244 (1993).
5. Wrachtrup, J. *et al.* *Nature* **363**, 244–245 (1993).
6. Praver, S. & Greentree, A. D. *Science* **320**, 1601–1602 (2008).
7. Childress, L. *et al.* *Science* **314**, 281–285 (2006).
8. Mamin, H. J., Poggio, M., Degen, C. L. & Rugar, D. *Nature Nanotech.* **2**, 301–306 (2007).

## NEUROSCIENCE

# Fragile dopamine

David Weinshenker and Stephen T. Warren

**Dopamine dysfunction, which is implicated in Parkinson's disease and drug addiction, seems an unlikely culprit in fragile X syndrome. A surprising set of findings means a rethink is required.**

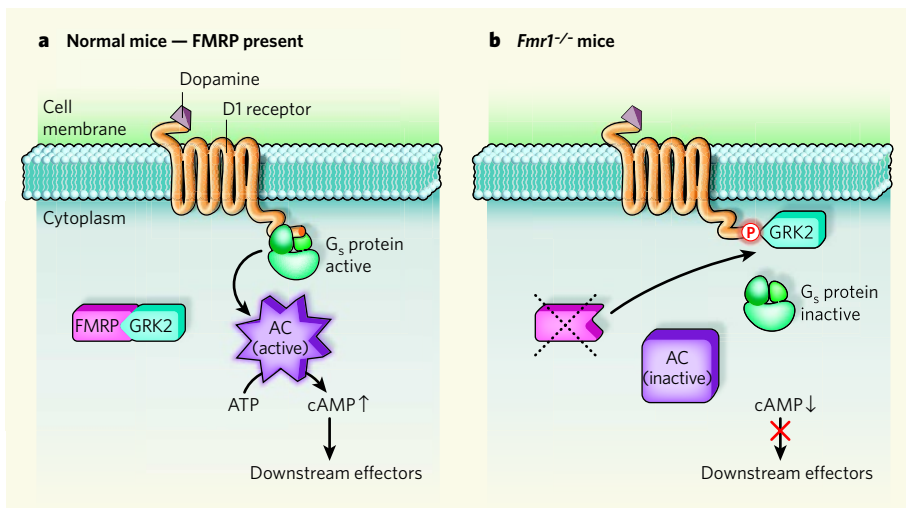
Fragile X syndrome is the commonest inherited form of mental retardation, with the patients often also having autism and attention-deficit hyperactivity disorder<sup>1</sup>. It is usually caused by the absence of the protein FMRP, which is encoded by *FMR1*, a gene on the X chromosome. Although FMRP function is not well understood, most studies concur that it is a selective RNA-binding protein that modulates the translation of its target messenger RNAs<sup>2</sup>. But this deceptively simple description of FMRP function omits any role for the neurotransmitter dopamine, despite the fact that some of the clinical and behavioural features of fragile X syndrome are reminiscent of dysfunction in dopamine-secreting neurons. Writing in *Neuron*, Wang *et al.*<sup>3</sup>

elucidate the role of FMRP in modulating dopamine signalling.

Classically, dopamine has been considered an essential mediator of behaviours such as reward-seeking and coordinated movement. Because these brain functions are not always impaired in fragile X syndrome, there has been little reason to link altered dopamine signalling to the aetiology or the manifestation of this disorder. Nevertheless, recent work<sup>4,5</sup> has expanded and refined dopamine's job description in the brain to encompass more cognitively relevant functions, such as involvement in reward-prediction error, motivation, ability to focus on pertinent environmental stimuli and goal-directed behaviours.

To explore a possible connection between





**Figure 1 | FMRP-mediated regulation of dopamine signalling.** **a**, Normally, FMRP binds to GRK2 and sequesters it in the cytoplasm, away from D1 dopamine receptors on the cell membrane. On dopamine binding, the D1 receptors couple to and activate G<sub>s</sub>, which in turn activates adenylate cyclase (AC). This enzyme then catalyses the formation of cAMP, which acts on various downstream effector molecules. **b**, Wang *et al.*<sup>3</sup> show that, in *Fmr1*<sup>-/-</sup> mice, which cannot produce FMRP (and presumably in patients with fragile X syndrome), GRK2 is no longer sequestered in the cytoplasm and is free to move to the cell membrane, where it phosphorylates D1 receptors. Consequently, even on dopamine binding, the phosphorylated D1 receptors have reduced affinity for G<sub>s</sub>. The signalling cascade stalls, cAMP levels remain low and effector molecules are no longer engaged.

dopamine signalling and FMRP activity, Wang *et al.*<sup>3</sup> studied a common experimental model of fragile X syndrome — mice lacking the *Fmr1* gene (*Fmr1*<sup>-/-</sup>) — and focused on the function of the D1 subtype of dopamine receptor in the brain's prefrontal cortex (PFC). The D1 receptor not only modulates the trafficking of the receptor for another neurotransmitter, glutamate, but also affects long-term potentiation, a form of synaptic plasticity thought to underlie learning and memory. The PFC is crucial for working memory, planning and attention<sup>6,7</sup>.

The authors show that, in the culture-grown PFC neurons of *Fmr1*<sup>-/-</sup> mice, cell-surface expression and phosphorylation of the GluR1 glutamate receptors, processes that are normally mediated by the D1 receptors, are reduced. Normally, D1 receptors are coupled to G<sub>s</sub> proteins, which stimulate the production of an intermediary signalling molecule called cyclic AMP by activating the enzyme adenylate cyclase (Fig. 1a). In *Fmr1*<sup>-/-</sup> mice, the ability of a D1-receptor agonist to raise cAMP levels in PFC neurons was attenuated. However, no deficit in cAMP production was detected in these mice when a direct activator of adenylate cyclase was used. These and other data indicate that FMRP is crucial for the coupling of D1 receptors to G<sub>s</sub> proteins.

Kinase enzymes known as GRKs can phosphorylate D1 receptors at several amino-acid residues, diminishing the receptors' ability to couple to G<sub>s</sub>. Wang *et al.*<sup>3</sup> show that, in PFC neurons from *Fmr1*<sup>-/-</sup> mice, the phosphorylation of D1 receptors is abnormally high. Moreover, one specific GRK — GRK2 — was particularly abundant in the cell-membrane fractions of these neurons, where functional D1 receptors are localized. The authors

hypothesized that, normally, FMRP interacts with GRK2 and prevents its activity. Hence, in the absence of FMRP in *Fmr1*<sup>-/-</sup> mice — and possibly in patients with fragile X syndrome — GRK2 is no longer regulated and becomes overactive (Fig. 1b).

Indeed, using imaging and biochemical approaches, the authors show that FMRP and GRK2 interact. Notably, these two proteins seem to associate only in the cytoplasm, suggesting that FMRP normally functions to sequester GRK2. In the absence of FMRP in *Fmr1*<sup>-/-</sup> mice, GRK2 seems to be free to move to the cell membrane, where it phosphorylates D1 receptors and disrupts their signalling. But what are the functional consequences of these deficits? Assessing the effects of a D1 agonist on long-term potentiation, Wang *et al.* find that, although stimulating D1 receptors enhances this form of plasticity in the PFC of normal mice, it has no effect on the corresponding brain region of *Fmr1*<sup>-/-</sup> animals.

Intriguing as these findings are, unanswered questions remain about the relevance of impaired dopamine signalling to the manifestation of fragile X syndrome. Wang and colleagues' finding that, following exposure to a D1 agonist cAMP production remains impaired in *Fmr1*<sup>-/-</sup> mice, is consistent with previous reports<sup>8,9</sup> of reduced cAMP levels and diminished cAMP responses in cells of patients with fragile X syndrome. But unlike the earlier studies, which attributed decreased cAMP levels to reduced adenylate-cyclase activity, Wang *et al.* observe normal cAMP responses to an activator of adenylate cyclase. Instead, they report that D1-receptor–G<sub>s</sub> coupling is impaired in *Fmr1*<sup>-/-</sup> mice. Clearly, these differences must be reconciled. Nonetheless, Wang and colleagues'

data will surely spark renewed interest in the role of cAMP in fragile X syndrome.

Another interesting aspect of the results<sup>3</sup> is the implication that the PFC is involved in fragile X syndrome. Patients with this disorder exhibit deficits in executive functions, showing inappropriate behaviours that are usually inhibited, and impairments<sup>10</sup> in working memory, cognitive flexibility and planning — functions attributed to the PFC. Moreover, these patients are often described as 'creatures of habit' by their caregivers and have difficulty adapting to even subtle changes in their daily routines<sup>11</sup>. In agreement with previous work, Wang and colleagues show that D1-receptor-mediated pathways are crucial for long-term potentiation and synaptic plasticity in the corticostriatal neural circuitry, which is believed to be important for learning new goal-directed behaviours and for the formation of habits.

But extended training can override the dopamine-dependence of actions and habits<sup>4</sup>. So one consequence of impaired dopamine signalling in fragile X syndrome could be that, once patients establish a routine reinforced by extended repetition and training, their corticostriatal circuits cannot adapt rapidly to change. Although the authors<sup>3</sup> demonstrate that a D1 agonist efficiently reverses the locomotor hyperactivity behaviour in *Fmr1*<sup>-/-</sup> mice, it will be crucial to establish whether abbreviated D1-receptor signalling contributes to other cognitively relevant characteristics.

In the 17 years since the identification of *FMR1*, progress towards an understanding of the mechanism behind fragile X syndrome has been swift, with recent attempts at pharmaceutical interventions<sup>12</sup> being particularly noteworthy. But Wang and co-workers' observation that deregulation of GRK2 activity in the absence of FMRP leads to impaired dopamine signalling is a reminder that we still have much to learn about the molecular, cognitive and behavioural manifestations of this disorder. ■

David Weinshenker and Stephen T. Warren

are in the Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia 30322, USA.

e-mails: dweinsh@emory.edu;

swarren@emory.edu

- Garber, K. B., Visootsak, J. & Warren, S. T. *Eur. J. Hum. Genet.* **16**, 666–672 (2008).
- Penagarikano, O., Mülle, J. G. & Warren, S. T. *Annu. Rev. Genomics Hum. Genet.* **8**, 109–129 (2007).
- Wang, H. *et al.* *Neuron* **59**, 634–647 (2008).
- Wickens, J. R., Horvitz, J. C., Costa, R. M. & Killcross, S. *J. Neurosci.* **27**, 8181–8183 (2007).
- Berridge, K. C. *Psychopharmacology (Berl.)* **191**, 391–431 (2007).
- Calabresi, P., Picconi, B., Tozzi, A. & Di Filippo, M. *Trends Neurosci.* **30**, 211–219 (2007).
- Brennan, A. R. & Arnsten, A. F. *Ann. NY Acad. Sci.* **1129**, 236–245 (2008).
- Berry-Kravis, E. & Huttenlocher, P. R. *Ann. Neurol.* **31**, 22–26 (1992).
- Kelley, D. J. *et al.* *Neurosci. Biobehav. Rev.* **32**, 1533–1543 (2008).
- Hooper, S. R. *et al.* *Neuropsychology* **22**, 36–47 (2008).
- Hatton, D. D. *et al.* *Am. J. Med. Genet. A* **108**, 105–116 (2002).
- Bassell, G. J. & Gross, C. *Nature Med.* **14**, 249–250 (2008).

# The changing face of HIV in China

HIV has advanced from high-risk groups such as intravenous drug users to some in the general population, according to comprehensive new data from the south of China. What needs to be done to halt its spread?

Lin Lu, Manhong Jia, Yanling Ma, Li Yang, Zhiwei Chen, David D. Ho, Yan Jiang and Linqi Zhang

The HIV-1/AIDS epidemic in China is at a critical juncture. Historically, HIV-1 infection has been largely confined to certain high-risk populations such as intravenous drug users and former blood and plasma donors in geographically disparate rural areas<sup>1-3</sup>. However, HIV-1 prevalence has now increased rapidly among men who have sex with men and among female sex workers<sup>4,5</sup>. It seems that China is following the path of some of the other Asian countries where HIV-1 infection is no longer confined to high-risk populations<sup>5</sup>.

Since the first cases among foreign tourists and local recipients of imported factor VIII in the mid 1980s, HIV-1 has spread to all of mainland China<sup>1,4</sup>. The current epidemic comprises largely of two affected populations: former blood and plasma donors in Henan and neighbouring provinces, and intravenous drug users in Yunnan and along drug-trafficking routes<sup>1,2</sup> (Fig. 1). Both of these populations stemmed from the infection of drug users from Yunnan's Dai and Jingpo ethnic minority groups in Yunnan in the late 1980s<sup>6,7</sup>.

Statistics from the Chinese Ministry of Health and UNAIDS have revealed a worrisome trend of the HIV-1 epidemic in China. As of October 2007, an estimated 700,000 infections had occurred<sup>4</sup>. Although the prevalence of infection remains low (0.04–0.07%), the new figure represents an 8% increase since 2006 (refs 4, 5). Remarkably, 38% of the cases were attributed to heterosexual contacts — more than triple the 11% in 2005 (ref. 4). In line with this trend, the proportion of women infected has doubled over the past decade<sup>4</sup>. As 90% of these women are of child-bearing age (15–44), this is likely to translate into more vertical transmission from mother to child<sup>4</sup>. Additionally, the proportion of cases among men who have sex with men increased eight-fold from 0.4% in 2005 to 3.3% in 2007 (ref. 4). These data suggest that the HIV-1 epidemic is expanding, and that more effective preventive measures are urgently needed.

## The epidemic in Yunnan

Located in southwest China, Yunnan has long been regarded as China's Shangri-la for its natural beauty. But now, with all of its 16 prefectures affected, it is a major site of



**Figure 1 | Pervasive spread.** The geographic distribution of cumulative reported HIV-1 infection in mainland China (source: ref. 4).

the AIDS epidemic. Yunnan's ethnic diversity is unrivalled in China, with 25 different ethnic minority groups representing one third of the province's population. Of these groups, 13 live along the border with Myanmar, Laos and Vietnam, and cross-border travel and commerce are common. Yunnan has a long history of opium/heroin trade, and the vast majority of illicit drugs in China are trafficked through Yunnan from the 'Golden triangle' of illicit opium production, encompassing Myanmar, Thailand, Laos and Vietnam (Fig. 1)<sup>8,9</sup>.

HIV-1 was detected in intravenous drug users in Yunnan in 1989 (ref. 10). It then also spread among other populations<sup>11</sup>. Between 1989 and 2006, 3.2 million blood samples were tested in Yunnan. This testing identified 48,951 HIV-1 cases, 3,935 AIDS patients, and 1,768 resultant deaths — representing about 25%, 8% and 13% of the national totals,

respectively. Prefectures bordering Myanmar and Vietnam were the first and the most severely affected.

Although the cumulative HIV-1 case load rose gradually from 1989 to 2003, there was a sharp rise in 2004, when 13,486 new cases were seen. This total is comparable to the number identified in the previous 16 years. Identification of these new cases was likely to be due to increased surveillance and testing since the estimated incidence rates remained relatively stable over time among the major risk groups. These estimates were determined by re-testing all seropositive samples from the surveillance effort using the BED assay<sup>12</sup> to detect those with low-affinity antibodies to HIV-1. Intravenous drug users had the highest incidence rate throughout the study, varying between 2.2% and 8.0% per year, whereas that for outpatients attending sexually transmitted infection (STI) clinics was 0.3–1.0%



per year and for pregnant women it was about 0.1% per year.

### Changing demographics

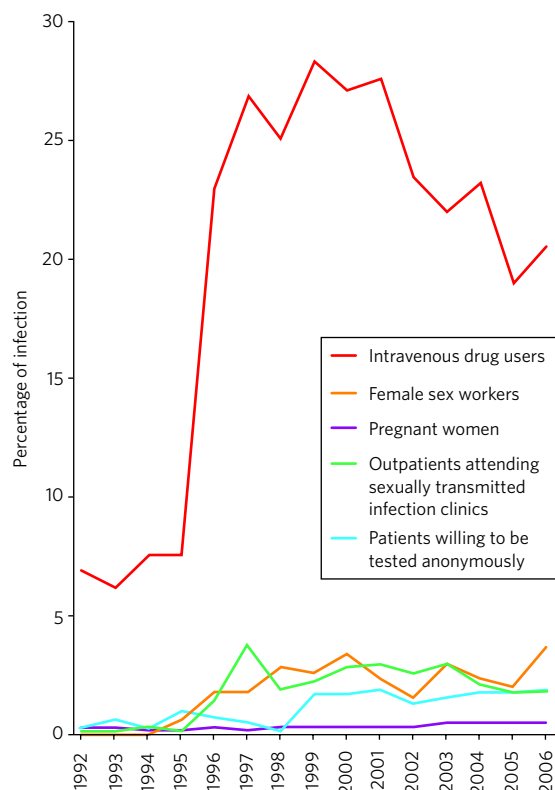
HIV-1 has hit different Yunnan populations disproportionately. Figure 2 shows provincial average HIV-1 prevalence rates over time among intravenous drug users, female sex workers, STI outpatients, pregnant women and an 'unlinked population' (patients admitted to the hospital who were willing to be tested for HIV-1 anonymously) based on the results from 97 sentinel surveillance sites located through the province. Within these groups, the highest prevalence rate has always been found in the intravenous drug users population. From 1992 to 1995, the average prevalence rate remained around 6%. In 1996, it jumped to about 22%, and then remained near that level. Prevalence rates among female sex workers and STI outpatient groups have been consistently lower, but follow a similar trend. The 'unlinked' and pregnant women populations have also experienced a similar pattern of prevalence increases, although the jumps occurred in 1999 and 2003, respectively. This sequential upsurge of infection among intravenous drug users, female sex workers, STI outpatients, then among the 'unlinked' and finally pregnant women, is reminiscent of what has been seen in other countries, where HIV-1 infection has spread from high-risk groups to some in the general population.

This trend in transmission mode is further illustrated by the fact that the proportion of cases among intravenous drug users has decreased from 100% in 1989 to 40% in 2006. Concurrently, heterosexual transmission has increased markedly, reaching 37.5% of infections in 2006.

Although most infections were in farmers from 1989 to 1995, more factory workers are now infected, and the number of infections among unemployed persons have come to rival those in farmers. In addition, whereas the Dai and Jingpo minorities were the most affected ethnic groups in 1989–95, Han Chinese overtook these minorities in 1996 and up to 2006 accounted for around 60% of infections.

Changes in age distribution are also evident. Although on average more than 95% of infected individuals have been aged 20–40, HIV-1 prevalence has increased among the 30–59 group and decreased among the 20–29 group. This could be attributed to ageing of infected individuals or to new infections of relatively older age groups over time. Nonetheless, high prevalence in the 20–29 and younger than 20 age groups suggests ongoing infection within the young population.

HIV-1 in Yunnan has also spread to the



**Figure 2 | Changing trends.** HIV-1 prevalence among various risk groups in Yunnan between 1992 and 2006.

female population. Before 1996, most infected individuals were male. However, from 1996–2006, the proportion of HIV-1-infected women gradually increased from 7.1% to 35%, and the male to female ratio decreased from 13:1 to 1.9:1.

### Virus evolution

With the dramatic changes in disease distribution, HIV-1 genetics in Yunnan have become increasingly complex. The initial HIV-1 epidemic among intravenous drug users in Yunnan in 1989 was caused by a mixture of viruses closely resembling European/North American subtype B and Thai subtype B (B')<sup>13</sup>. But by 1996, the B' subtype began to dominate<sup>13,14</sup>. During the same period, a second epidemic took root among intravenous drug users in Yunnan, with strains genetically related to subtype C viruses from India<sup>15</sup>. Co-existence of multiple subtypes led to the formation of circulating recombinant forms (CRFs) of HIV-1 — CRF07\_BC and CRF08\_BC among intravenous drug users along drug trafficking routes<sup>16</sup> and CRF01\_AE in Chinese sex workers who had worked in Thailand<sup>17</sup>. In the mid-1990s, viruses closely related to CRF01\_AE and CRF08\_BC in Yunnan were identified among intravenous drug users in Guangxi<sup>18</sup>. Further novel recombinants arose in subsequent years<sup>19</sup>.

We compared over 500 nucleotide sequences from Yunnan<sup>6,7</sup> with those from other provinces in China and neighbouring countries. Comparing sequences from the HIV-1 gag p17

gene with reference sequences from the HIV-1 Database ([www.hiv.lanl.gov/content/index](http://www.hiv.lanl.gov/content/index)), we identified the three main subtypes of HIV-1 found in Yunnan. These subtypes are those clustering closely with subtype C, CRF07\_BC, or CRF08\_BC (53.0%); those with CRF01\_AE or CRF15\_01B (40.5%); and those with subtype B (6.5%). Notably, more than 90% of infected intravenous drug users had C/CRF07\_BC/CRF08\_BC viruses, whereas 85.4% of CRF01\_AE/CRF15\_01B infections were acquired through sexual transmission. Furthermore, sequences in the C/CRF07\_BC/CRF08\_BC group were found throughout Yunnan, while those in the CRF01\_AE/CRF15\_01B group were largely confined to prefectures bordering Myanmar. Sequences in the subtype B groups have only been identified in the Dehong and Baoshan prefectures in Yunnan.

The dominant C/CRF07\_BC/CRF08\_BC viruses in Yunnan are related to strains in Guangxi province and distant Xinjiang province, supporting the notion that HIV-1 has spread along known drug-trafficking routes. In contrast, sequences similar to CRF01\_AE/CRF15\_01B have been largely confined to Yunnan and are closely related to strains from Thailand, Myanmar and Vietnam. Subtype B sequences from Yunnan are genetically similar to those from former blood donors in Henan and adjacent provinces, and can be broadly classified into two major groups: one with sequences similar to those from Thailand and Myanmar, and the other with sequences more similar to those in France and the United States. These results are consistent with the hypothesis that HIV-1 spread from Yunnan to central China, and suggest multiple introductions of HIV-1 from foreign countries to Yunnan<sup>6</sup>.

### Challenge and opportunity

Over the past 20 years, HIV-1 in Yunnan has overcome preventive measures to spread beyond high-risk populations. The dramatic increase in sexual transmission has changed the demographic profile of those infected. As the epidemic continues to expand, the genetic makeup of HIV-1 subtypes have become increasingly complex, potentially posing greater challenges to our efforts in antiretroviral treatment and vaccine development.

In light of the observed demographic changes, HIV-1 prevention strategies must focus more on stopping sexual transmission of HIV-1 within high-risk groups and halting the spread to the general public. There are urgent needs to scale up and integrate those proven successful prevention programmes such as condom promotion among female sex workers; drug rehabilitation, needle exchange

and methadone maintenance for intravenous drug users; and free antiretroviral therapy for those infected.

There is an old Chinese saying: “When there is a crisis, there is an opportunity.” Indeed, as HIV-1 plagues certain high-risk groups in China, there is still a window of opportunity to prevent further spread to the general population. The time to act is now. ■

1. Wu, Z., Sullivan, S. G., Wang, Y., Rotheram-Borus, M. J. & Detels, R. *Lancet* **369**, 679–690 (2007).
2. Wu, Z. Y., Liu, Z. Y. & Detels, R. *Lancet* **346**, 61–62 (1995).
3. Kaufman, J. & Jing, J. *Science* **296**, 2339–2340 (2002).
4. State Council AIDS Working Committee Office and UN Theme Group on AIDS in China, *A Joint Assessment of HIV/AIDS Prevention, Treatment and Care in China* (2007).

5. Gill, B., Huang, Y. & Lu, X. *Demography of HIV/AIDS in China* Center for Strategic and International Studies (2007).
6. Zhang, L. *et al. J. Virol.* **78**, 13591–13599 (2004).
7. Zhang, Y. *et al. PLoS Med.* **3**, e443 (2006).
8. US Department of State Bureau for International Narcotics and Law Enforcement Affairs. *International Narcotics Control Strategy Report Volume I Drug and Chemical Control* (2007).
9. Chin, K.-L. & Zhang, S. US Department of Justice Report, *The Chinese Connection: Cross-border Drug Trafficking between Myanmar and China* (2007).
10. Zheng, X. *et al. AIDS* **8**, 1141–1147 (1994).
11. Xiao, Y., Kristensen, S., Sun, J., Lu, L. & Vermund, S. H. *Soc. Sci. Med.* **64**, 665–675 (2007).
12. Parekh, B. S., Pau, C. P., Kennedy, M. S., Dobbs, T. L. & McDougal, J. S. *AIDS Res. Hum. Retroviruses* **17**, 137–146 (2001).
13. Graf, M. *et al. AIDS Res. Hum. Retroviruses* **14**, 285–288 (1998).
14. Weniger, B. G., Takebe, Y., Ou, C. Y. & Yamazaki, S. *AIDS* **8**, S13–S28 (1994).

8, S13–S28 (1994).

15. Luo, C. C. *et al. Lancet* **345**, 1051–1052 (1995).
16. Su, L. *et al. J. Virol.* **74**, 11367–11376 (2000).
17. Cheng, H., Zhang, J., Capizzi, J., Young, N. L. & Mastro, T. D. *Lancet* **344**, 953–954 (1994).
18. Piyasirisilp, S. *et al. J. Virol.* **74**, 11286–11295 (2000).
19. Yang, R. *et al. J. Virol.* **77**, 685–695 (2003).

**Acknowledgement** This work is supported by National Basic Research Program (also called 973 Program) of Chinese Ministry of Science and Technology to L.Z. (2006CB504200) and by the Tenth Five-year Key Technologies R&D Programme of China: 2004BA719A14-1, 2004BA719A14-2. We also thank Mark Goldin for helpful suggestions.

**Author information** Correspondence should be addressed to L.Z. (e-mail lzhang07@hotmail.com or jiangyan03@263.net). L.L. and M.J. contributed equally to this work.

## Authorship

Lin Lu<sup>1</sup>, Manhong Jia<sup>1</sup>, Yanling Ma<sup>1</sup>, Li Yang<sup>1</sup>, Zhiwei Chen<sup>2</sup>, David D. Ho<sup>3,4</sup>, Yan Jiang<sup>5\*</sup> and Linqi Zhang<sup>4,6\*</sup>

<sup>1</sup>Yunnan Center for Disease Control and Prevention, Yunnan, People's Republic of China. <sup>2</sup>AIDS Institute, The University of Hong Kong Li Ka Shing Faculty of Medicine, Hong Kong SAR, People's Republic of China. <sup>3</sup>Aaron Diamond AIDS Research Center, The Rockefeller University, New York, USA. <sup>4</sup>Comprehensive AIDS Research Center, Tsinghua University, Beijing, People's Republic of China. <sup>5</sup>National AIDS Reference Laboratory, National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, People's Republic of China. <sup>6</sup>AIDS Research Center, State Key Laboratory for Molecular Virology and Genetic Engineering, Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, People's Republic of China.



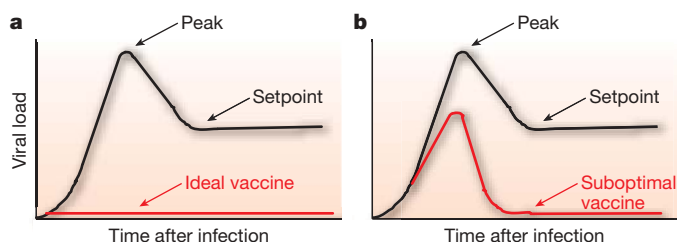
# Challenges in the development of an HIV-1 vaccine

Dan H. Barouch<sup>1</sup>

**The development of a safe and effective human immunodeficiency virus (HIV)-1 vaccine is a critically important global health priority. Despite recent advances in our understanding of HIV-1 pathogenesis and immunology, however, major scientific obstacles remain. Prototype HIV-1 vaccine candidates aimed at eliciting humoral and cellular immune responses have so far failed to protect against HIV-1 infection or to reduce viral loads after infection in clinical efficacy studies. A renewed and coordinated commitment to basic discovery research, preclinical studies and clinical trials will therefore be required to overcome the hurdles currently facing the field. Here I review key challenges and future prospects in the quest to develop a prophylactic HIV-1 vaccine.**

It has been 25 years since HIV-1 was identified as the causative agent for AIDS<sup>1–5</sup>. More than 60 million people worldwide have been infected with HIV-1, mostly in the developing world, and nearly half of these individuals have died. The development of a safe and effective HIV-1 vaccine would undoubtedly be the best solution for the ultimate control of the worldwide AIDS pandemic<sup>6</sup>, but unfortunately HIV-1 vaccine development efforts have not yet proven successful. The extraordinary diversity of HIV-1, the capacity of the virus to evade adaptive immune responses, the inability to induce broadly reactive antibody responses, the early establishment of latent viral reservoirs, and the lack of clear immune correlates of protection represent unprecedented challenges for vaccine development.

The goal of an HIV-1 vaccine would be either to prevent infection or to reduce viral loads and clinical disease progression after infection (Fig. 1). An ideal vaccine would completely block infection and provide sterilizing immunity. Although such a vaccine would be optimal, this degree of protection is not even achieved with most clinically licensed vaccines. In contrast, most licensed viral vaccines seem to function by controlling subclinical viral replication and by preventing clinical disease. It may therefore be more realistic to develop a suboptimal HIV-1 vaccine that fails to prevent infection but that provides partial immune control of viral replication after infection. Such partial control, as exemplified by a reduction in peak and setpoint viral loads after infection, has been demonstrated in certain preclinical studies by vaccines that elicit T lymphocyte responses.



**Figure 1 | Goals of an HIV-1 vaccine.** After infection, HIV-1 replicates exponentially to a peak level and then is partially controlled to a viral setpoint level (black). **a**, An ideal vaccine would protect against infection and afford sterilizing immunity (red). **b**, A suboptimal vaccine would result in decreased peak and setpoint viral loads after infection (red).

Moreover, because viral loads represent a principal determinant of HIV-1 transmission<sup>7</sup>, it is conceivable that such a partially protective vaccine might have substantial impact on a population level.

Despite the urgent need for an HIV-1 vaccine, only two vaccine concepts have completed clinical efficacy studies so far. The first vaccine concept used monomeric HIV-1 Env gp120 protein, and the aim of this strategy was to induce Env-specific humoral immune responses. In early-phase clinical trials, gp120 immunogens elicited type-specific binding antibodies but failed to induce broadly reactive neutralizing antibodies<sup>8,9</sup>. In two phase 3 efficacy trials sponsored by the biotechnology company VaxGen, these vaccine candidates afforded no detectable protective efficacy<sup>10,11</sup>, indicating that these type-specific antibody responses were insufficient to protect against HIV-1 infection in humans. Another phase 3 study evaluating the efficacy of a recombinant canarypox vector prime/gp120 protein boost vaccine regimen is currently underway. The second vaccine concept that has completed clinical efficacy studies involved replication-incompetent recombinant adenovirus serotype 5 (rAd5) vectors expressing HIV-1 Gag, Pol and Nef. The aim of this strategy was to elicit HIV-1-specific cellular immune responses. Early-phase clinical trials demonstrated that rAd5 vector-based vaccines elicited cellular immune responses in most subjects, although these responses were partially suppressed in individuals with pre-existing Ad5-specific neutralizing antibodies<sup>12</sup>. Phase 2b efficacy trials sponsored by Merck and the National Institutes of Health (NIH) were unexpectedly terminated when the first planned interim analysis showed that this vaccine failed to protect against infection or to reduce viral loads after infection, and that vaccinees with pre-existing Ad5-specific neutralizing antibodies exhibited an enhanced rate of HIV-1 acquisition<sup>13</sup>. These results have highlighted new scientific challenges and have led to substantial debate regarding the optimal path forward for the HIV-1 vaccine field.

## Virologic and immunologic challenges

The challenges in the development of a prophylactic HIV-1 vaccine are unprecedented (Box 1). The extraordinary worldwide diversity of HIV-1 presents perhaps the greatest hurdle<sup>14</sup>. Driven by the error-prone reverse transcriptase, the HIV-1 M group has diversified into nine divergent clades as well as multiple circulating recombinant forms. Amino acid sequences of Env can differ up to 20% within a particular clade and over 35% between clades<sup>14,15</sup>. A vaccine

<sup>1</sup>Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA.

**Box 1 | Challenges in the development of a prophylactic HIV-1 vaccine**

- (1) Extensive viral clade and sequence diversity.
- (2) Early establishment of latent viral reservoirs.
- (3) Immune correlates of protection unclear.
- (4) Viral evasion of humoral and cellular immune responses.
- (5) Antibody responses typically type-specific.
- (6) No method exists to elicit broadly reactive neutralizing antibodies.
- (7) Attenuated viruses unsafe for human use.
- (8) Lack of a small-animal model.
- (9) Little pharmaceutical interest.

immunogen will therefore need to contend with a remarkably high degree of viral diversity, and vaccine protection will necessarily be dependent on the capacity of immune responses to cross-react with highly heterologous viruses. Although cross-reactive humoral and cellular immune responses against conserved regions of the virus have been reported, it is reasonable to assume that protective efficacy will diminish substantially with increasing divergence between vaccine antigens and infecting viruses.

Another key challenge is the lack of clear immune correlates of protection in humans, because HIV-1-infected patients are unable to eradicate the virus. Suggestive evidence regarding immune correlates of protection might be obtained from viral challenge studies in non-human primates and from studies of HIV-1-infected individuals who spontaneously control viral replication to very low levels. However, definitive immune correlates of protection will probably only emerge in the context of successful vaccine efficacy studies in humans.

**HIV-1-specific humoral immunity.** Virus-specific neutralizing antibody titres represent key immune correlates of protection for most licensed viral vaccines, and thus early studies focused on developing HIV-1 Env subunit immunogens. Advances in our understanding of Env structure and function have begun to elucidate why generating broadly reactive neutralizing antibodies to HIV-1 by vaccination may be so difficult<sup>16</sup>. The HIV-1 Env glycoprotein is a trimer on the virion surface with extensive N-linked glycosylation that effectively shields many conserved epitopes from antibody recognition<sup>17,18</sup>. Highly immunogenic variable loops also elicit type-specific antibodies that may redirect humoral responses away from conserved regions. In addition, key conserved regions, such as the binding site of the chemokine co-receptor, are only formed after Env binds its cellular receptor CD4 and undergoes an extensive conformational change<sup>19</sup>. The development of mutations in N-linked glycans has also been shown to lead to rapid evasion of host neutralizing antibody responses<sup>20,21</sup>.

Nevertheless, broadly reactive neutralizing antibody activity has been identified in a small number of HIV-1-infected subjects, and this reactivity seems to be largely directed against conserved regions of the Env glycoprotein such as the CD4-binding site<sup>22</sup>. The broadly reactive monoclonal antibody b12 also binds to the CD4-binding site, suggesting that this region of Env may represent a critical point of vulnerability that is potentially amenable to neutralization<sup>23</sup>. However, the CD4-binding site is recessed and only partially accessible to antibody binding. Another conserved region is the membrane-proximal external region (MPER) of gp41, which represents the target of the broadly reactive monoclonal antibodies 2F5 and 4E10. However, MPER-specific neutralizing antibodies may be difficult to elicit by vaccination for multiple reasons, including tolerance control and immunoregulation<sup>24</sup>, sequestration of the epitope in the lipid membrane<sup>25</sup>, exposure of the epitope only transiently during viral entry<sup>26</sup>, or possibly a combination of multiple factors.

The development of immunogens that induce broadly reactive neutralizing antibodies is perhaps the most important priority for the HIV-1 vaccine field<sup>16</sup>. Proof-of-concept passive transfer studies in non-human primates have shown that administration of high doses of broadly reactive monoclonal antibodies can afford sterilizing protection from infection, thus demonstrating the potential of

virus-specific humoral immunity<sup>27,28</sup>. However, it has not been possible to induce such broadly reactive neutralizing antibodies by vaccination so far. Although there has been substantial progress in our understanding of Env structure and function, there are currently no vaccine candidates that are aimed at eliciting broadly reactive Env-specific neutralizing antibodies in clinical trials. It is likely that next-generation Env immunogens will need to be engineered antigens. Strategies that are being pursued include generating biochemically stabilized Env trimers, constraining Env immunogens in structurally defined conformations, scaffolding conserved neutralization epitopes onto foreign proteins, developing methods to circumvent immunoregulation, and designing immunogens to target specific regions such as the CD4-binding site, the MPER region and structurally conserved elements of the V3 loop. The relevance of non-neutralizing antibodies that mediate other effector functions such as antibody-dependent cell-mediated virus inhibition, complement activation and phagocytosis is also being investigated.

**HIV-1-specific cellular immunity.** Virus-specific T lymphocyte responses are believed to have a critical role in controlling HIV-1 replication and are therefore being actively explored in vaccine development strategies. Early studies showed that virus-specific CD8<sup>+</sup> T lymphocyte responses emerge during acute infection coincident with initial control of primary viremia<sup>29–31</sup>. Potent cellular immune responses have also been reported in long-term non-progressors<sup>32</sup>, and specific HLA alleles and the breadth of Gag-specific T lymphocyte responses have been correlated with control of viral replication in HIV-1-infected individuals<sup>33,34</sup>. These data indicate the potential importance of cellular immune responses in immune control of HIV-1. Concordant with these observations, experimental depletion of CD8<sup>+</sup> lymphocytes has been shown to abrogate immune control of simian immunodeficiency virus (SIV) replication in rhesus monkeys<sup>35,36</sup>.

A limitation of virus-specific T lymphocyte responses is the propensity of the virus to accumulate mutations in T lymphocyte epitopes and to evade cellular immune control<sup>37–39</sup>. It is therefore likely that the breadth of epitope-specific T lymphocyte responses will prove critical for an HIV-1 vaccine, not only to maximize immunologic coverage of HIV-1 diversity but also to minimize the potential for viral escape from recognition by T lymphocytes. However, the breadth of vaccine-elicited cellular immune responses may be limited by immunodominance constraints and by the inherent tendency of CD8<sup>+</sup> T lymphocyte responses to be highly focused on a limited number of epitopes.

Recent advances in the characterization of T lymphocyte responses by multiparameter flow cytometry have highlighted the functional diversity of virus-specific T lymphocytes in terms of cytokine secretion, degranulation, proliferation and other effector functions in various subpopulations of effector and memory T lymphocytes. It is likely that the complex functionality of T lymphocytes may ultimately prove more relevant than interferon- $\gamma$  secretion as measured by enzyme-linked immunospot (ELISPOT) assays for the evaluation of vaccine-elicited cellular immune responses. Polyfunctional T lymphocytes capable of performing multiple functions have been reported in long-term non-progressors<sup>40</sup>, in recipients of effective vaccines such as vaccinia<sup>41</sup>, and in certain preclinical challenge studies<sup>42</sup>. These considerations suggest that the breadth<sup>43</sup> and quality<sup>44</sup> of T lymphocyte responses may prove critical in addition to the magnitude of these responses.

Perhaps the most significant limitation of vaccine-elicited cellular immune responses is that they will probably not protect against acquisition of HIV-1 infection. As a result, vaccine-induced T lymphocyte responses will presumably be unable to prevent lifelong infection, because the virus rapidly establishes latent reservoirs<sup>45,46</sup>. Moreover, it is unclear whether vaccine-elicited T lymphocytes will be able to function rapidly enough given that important immunopathologic events occur within the first few days of acute HIV-1 infection. HIV-1 preferentially infects HIV-1-specific CD4<sup>+</sup> T



lymphocytes<sup>47</sup> and rapidly depletes most memory CD4<sup>+</sup> T lymphocytes in gut-associated lymphoid tissue within the first 4–10 days of infection<sup>48–50</sup>. This sets the stage for progressive immunodeficiency as well as for chronic immune activation, which probably results at least in part from microbial translocation across damaged gastrointestinal mucosa<sup>51</sup>. Given the time required for vaccine-induced CD8<sup>+</sup> T lymphocyte responses to expand after infection, it may be difficult for vaccine-elicited T lymphocytes to prevent these early immunopathologic events completely<sup>52</sup>.

### Current HIV-1 vaccine strategies

**Traditional strategies.** Vaccine strategies for HIV-1 can be divided into traditional and novel vaccine approaches (Box 2). Traditional vaccine technologies include live attenuated viruses, whole killed viruses and protein subunits. Although these approaches have proven enormously successful for the development of vaccines against other viruses, they all have substantial limitations in terms of their utility for HIV-1. Live attenuated viruses have afforded substantial protective efficacy against SIV challenges in rhesus monkeys<sup>53,54</sup>, but they are unlikely to be used in humans owing to significant safety concerns<sup>55–57</sup>. In contrast, whole killed viruses<sup>58</sup> and protein subunits<sup>10,11</sup> are limited by their inability to induce broadly reactive neutralizing antibody responses as well as by their inability to elicit CD8<sup>+</sup> T lymphocyte responses. Recent data, however, suggest that Toll-like receptor adjuvants may increase the utility of protein subunit immunogens<sup>59,60</sup>.

**Novel strategies.** New vaccine strategies include gene-delivery technologies such as plasmid DNA vaccines and live recombinant vectors that are engineered to express HIV-1 antigens. Plasmid DNA vaccines offer considerable promise in terms of simplicity and versatility, but multiple injections of high doses of DNA vaccines are typically required to elicit detectable immune responses in non-human primates and humans<sup>61,62</sup>. Substantial research is therefore focused on the development of adjuvants for DNA vaccines<sup>63,64</sup> and improved delivery technologies such as *in vivo* electroporation<sup>65,66</sup>. Recombinant vectors include attenuated or replication-incompetent viruses, most notably adenoviruses<sup>12,67,68</sup> and poxviruses<sup>69,70</sup>. Viral vectors, administered either alone or in the context of heterologous DNA prime/vector boost regimens, represent most HIV-1 vaccine candidates that are currently in clinical trials. Other viral vectors that are being evaluated include vesicular stomatitis virus, adeno-associated virus, Venezuelan equine encephalitis virus, cytomegalovirus, herpes simplex virus and measles virus. Bacterial and mycobacterial vectors are also being explored, including *Salmonella*, *Listeria* and Bacille Calmette-Guérin (BCG).

### The STEP study

**Preclinical background.** Recombinant Ad5 vectors were selected for development by Merck on the basis of preclinical vector comparison studies that showed that rAd5 vectors were more immunogenic than multiple other vector modalities in rhesus monkeys<sup>67,71</sup>. Moreover, rAd5 vectors expressing SIV Gag afforded marked reductions of viral loads after challenge of rhesus monkeys with the chimaeric simian–human immunodeficiency virus (SHIV)-89.6P (ref. 67). However, it was also observed that the same vaccine afforded minimal to no control of peak or setpoint viral loads after challenge with SIV<sub>MAC239</sub> (ref. 72), indicating that SIV challenges were considerably more stringent than SHIV-89.6P challenges.

A DNA prime/rAd5 boost regimen expressing SIV Gag afforded a brief (90 days) and marginal (0.8 log) reduction of peak viral loads after SIV<sub>MAC239</sub> challenge<sup>72</sup>, but this effect was only observed in rhesus monkeys that were selected to express the major histocompatibility complex (MHC) class I molecule Mamu-A\*01, which is associated with efficient virologic control<sup>73–75</sup>. A DNA prime/rAd5 boost regimen expressing multiple SIV antigens afforded increased protective efficacy in Mamu-A\*01-positive rhesus monkeys<sup>76</sup>, indicating that expanding the breadth of cellular immune responses improves protection. However, neither rAd5 alone nor DNA prime/rAd5 boost regimens have been able to reduce setpoint viral loads after SIV challenge of Mamu-A\*01-negative rhesus monkeys so far<sup>72,77</sup>.

**Clinical studies.** The Merck HIV-1 vaccine candidate was formulated as a trivalent mixture of rAd5 vectors expressing HIV-1 clade B Gag, Pol and Nef. Phase 1 clinical trials suggested that this vaccine was generally well tolerated and immunogenic in most volunteers<sup>12</sup>. However, as predicted by preclinical studies<sup>61</sup>, responses to this vaccine were partially suppressed in individuals with pre-existing neutralizing antibodies against the vaccine vector. Because 30–40% of individuals in the United States and Western Europe and 80–90% of people in sub-Saharan Africa have pre-existing Ad5-specific neutralizing antibodies<sup>78–81</sup>, the impact of anti-vector immunity was predicted to be a limitation of rAd5 vectors.

Two phase 2b ‘proof-of-concept’ efficacy studies were initiated by Merck and the National Institutes of Health to determine whether HIV-1-specific cellular immune responses induced by this vaccine regimen would prevent HIV-1 infection or would reduce viral loads after infection. HIV Vaccine Trials Network (HVTN) 502, also known as the ‘STEP’ study, was a 3,000-subject study in the Americas, the Caribbean and Australia. HVTN 503, also called ‘Phambili’ (which means ‘to move forward’ in Xhosa), was designed as a parallel 3,000-subject study in South Africa.

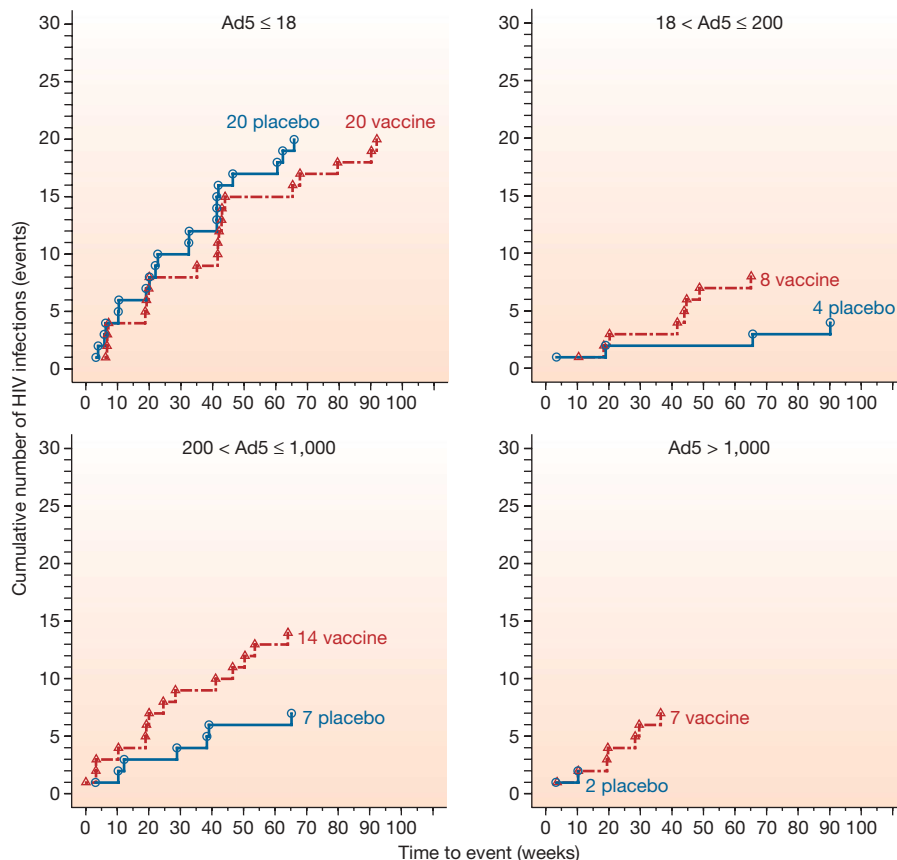
On 18 September 2007, HVTN 502 was unexpectedly terminated at the first planned interim analysis when the Data and Safety Monitoring Board declared futility in the study achieving its primary end points<sup>13</sup>. Moreover, in subjects with pre-existing Ad5-specific neutralizing antibody titres, a greater number of HIV-1 infections occurred in vaccinees than in placebo recipients (Fig. 2). Although the biological basis for this observation remains unclear, these data suggest that vaccination with rAd5 vectors may be associated with an increased risk of HIV-1 acquisition in this subgroup. Post-hoc multivariate analysis further suggested that the greatest increased risk was in men who had pre-existing Ad5-specific neutralizing antibodies and who were uncircumcised.

It is currently unclear whether the lack of efficacy in the STEP study simply represents the failure of the Merck rAd5-Gag/Pol/Nef vaccine product or whether this might be the harbinger of the failure of the T-cell vaccine concept overall. It is likely that substantial data will emerge from detailed immunologic analyses of vaccinees who subsequently became infected, and it is possible that the rAd5-Gag/Pol/Nef vaccine failed to induce sufficient magnitude, breadth or quality of cellular immune responses<sup>82</sup>. At the present time, therefore, it would seem premature to consider the failure of this single study as the failure of T-cell-based vaccines in general.

The apparent increased risk of HIV-1 acquisition in vaccinees with pre-existing Ad5-specific neutralizing antibodies was unexpected, and this finding highlights our lack of understanding of the parameters that determine susceptibility to HIV-1 infection. The biological basis for this observation remains unclear. One hypothesis is that rAd5 vaccination of individuals with pre-existing Ad5-specific neutralizing antibodies may have resulted in potent anamnestic Ad5-specific CD4<sup>+</sup> T lymphocytes that were increased targets for HIV-1 infection. However, early data have suggested that Ad5-specific T lymphocyte responses after rAd5 vaccination are actually lower in individuals with pre-existing Ad5-specific neutralizing antibodies than in those without pre-existing Ad5-specific neutralizing antibodies (J. McElrath,

#### Box 2 | Current HIV vaccine strategies

- (1) Traditional strategies
  - Live attenuated viruses
  - Whole killed viruses
  - Protein subunits
- (2) Novel strategies
  - Plasmid DNA vaccines
  - Live recombinant vectors



**Figure 2 | Cumulative HIV-1 infections in men enrolled in the STEP study stratified by pre-existing Ad5-specific neutralizing antibody titre.** Cumulative infections as of 17 October 2007 in men enrolled in the STEP study (HVTN 502) evaluating the Merck rAd5-Gag/Pol/Nef vaccine are

depicted. Infections in vaccinees (red) and placebos (blue) are shown in individuals stratified by their pre-existing Ad5-specific neutralizing antibody titres. Data represent the modified intent-to-treat population. Image courtesy of M. Robertson, Merck Research Laboratories.

unpublished observations). An alternative hypothesis is that Ad5-specific neutralizing antibodies may have opsonized rAd5 vectors after immunization, resulting in altered tropism or inflammatory responses. It is also possible that pre-existing Ad5-specific neutralizing antibodies may have been a marker for other confounding variables that have not yet been identified.

### A STEP forward?

Despite the disappointing results of the STEP study, several key lessons have already been learned. First, it is clear that the path forward towards an HIV-1 vaccine will be neither simple nor straightforward. Second, the importance of understanding both systemic and mucosal immune responses to vaccine vectors is paramount. Third, the biological determinants of HIV-1 acquisition and the impact that vector-specific and antigen-specific mucosal immune responses may have on this process will require intensive investigation. Fourth, clinical vaccine studies will need to adapt to the safety concerns raised by the STEP study, such as possibly excluding subjects who have pre-existing neutralizing antibodies to the vaccine vector that is used until this phenomenon is more completely understood. Fifth, future T-cell-based vaccine candidates should be prioritized for clinical efficacy studies only if they are convincingly superior to the homologous rAd5-Gag/Pol/Nef regimen that has failed. Sixth, non-human primate challenge models should be recalibrated on the basis of the STEP study to guide future HIV-1 vaccine development.

The protection afforded by the homologous rAd5 regimen against SHIV-89.6P indicates that this model lacks sufficient stringency for the evaluation of T-cell-based vaccine candidates. Although the more stringent SIV challenge model cannot be considered to be validated until there is a successful clinical efficacy study in humans, it seems

reasonable to use SIV<sub>MAC239</sub> or SIV<sub>MAC251</sub> as challenge viruses for evaluating next-generation vaccine candidates (Box 3). Preclinical challenge studies need to be adequately powered with sufficient follow-up time, and the vaccine schedule and dose should model the proposed clinical regimen. For optimal stringency, studies should exclude rhesus monkeys that express MHC class I alleles that are specifically associated with efficient virologic control, such as Mamu-A\*01, Mamu-B\*17 and Mamu-B\*08. The use of homologous Env antigens that may inappropriately overestimate protective efficacy should also be avoided. Mucosal challenges may offer certain physiological advantages over intravenous challenges, and these challenge models should therefore be developed. Finally, increased emphasis should be placed on assessing the capacity of promising vaccine candidates to protect against highly heterologous SIV challenges, because infecting viruses in humans will almost certainly be heterologous to any vaccine sequence. Because very few heterologous SIV challenge studies have been completed so far, a practical approach may be to determine the protective efficacy of promising vaccine

### Box 3 | Recommendations for preclinical challenge studies of T-cell-based vaccines

- (1) Use stringent challenge virus (SIV<sub>MAC239</sub>, SIV<sub>MAC251</sub>).
- (2) Design study with adequate power and follow-up time.
- (3) Model clinical regimen with vaccine schedule and dose.
- (4) Select rhesus monkeys that lack MHC alleles associated with efficient virologic control (Mamu-A\*01, Mamu-B\*17, Mamu-B\*08).
- (5) Avoid the use of a homologous Env antigen.
- (6) Assess promising vaccine concepts against both homologous and heterologous viral challenges.



candidates against both homologous and heterologous SIV challenges. It is currently debated whether non-human primate challenge studies should be used as a formal 'gatekeeper' for advancing vaccine candidates into clinical efficacy studies, because the capacity of this model to predict the results of clinical efficacy studies remains unclear. Nevertheless, it would seem reasonable to give a relative priority to the development of vaccine candidates that lead to durable control of setpoint viral loads after SIV<sub>MAC239</sub> or SIV<sub>MAC251</sub> challenge.

The STEP study has also had a major impact on other HIV-1 vaccine programmes in the field. HVTN 503 was terminated as it used the same rAd5-based vaccine candidate that was used in HVTN 502. The NIH Vaccine Research Center has developed a DNA prime/rAd5 boost vaccine regimen expressing clade B Gag-Pol and multiclade Env antigens. This vaccine candidate has been shown to be immunogenic in most individuals in phase 1 studies, particularly for the Env antigens<sup>62,68,83</sup>. In preclinical studies, a DNA prime/rAd5 boost vaccine regimen expressing SIV Gag, Pol, Nef and Env antigens afforded a 1.1 log reduction of peak viral loads for 112 days after a homologous SIV<sub>MAC251</sub> challenge<sup>77</sup>. However, no durable control of setpoint viral loads was observed with this vaccine, although delayed progression to AIDS-related mortality was evident<sup>77</sup>. NIH recently announced that it will not proceed with a large phase 2b efficacy study known as PAVE 100, although a smaller, more focused efficacy study with this vaccine candidate is still under consideration<sup>84</sup>. DNA prime/poxvirus boost regimens are also being evaluated using modified vaccinia Ankara (MVA)<sup>69</sup> and NYVAC<sup>70</sup> vectors, and phase 1 clinical trials have demonstrated immunogenicity in most volunteers. Central to all of these programmes, however, is the hypothesis that DNA priming before vector boosting will improve protective efficacy. This has been observed in some<sup>72</sup> but not all<sup>77</sup> SIV challenge studies, and thus it still remains an open question that requires further investigation and should be considered a high priority.

New rAd vectors derived from Ad serotypes that are rare in human populations are also being explored as a strategy to evade pre-existing Ad5-specific neutralizing antibodies. It is hoped that such vectors may offer immunologic as well as safety advantages as compared with rAd5 vectors by circumventing pre-existing vector-specific neutralizing antibodies. However, these possibilities have not yet been confirmed in clinical trials. Current strategies include the development of rare serotype rAd26, rAd35 and rAd48 vectors<sup>78,79,85</sup>; chimaeric rAd5HVR48 vectors in which dominant Ad5-specific neutralizing antibody epitopes have been exchanged<sup>86</sup>; and non-human rAd vectors<sup>87,88</sup>. Rare serotype rAd vectors are biologically different from rAd5 vectors in terms of their cellular receptors, tropism, intracellular trafficking pathways and innate immune profiles. Moreover, rAd26 and rAd48 vectors have been shown to elicit T lymphocyte responses of a substantially different phenotype as compared with rAd5 vectors<sup>89</sup>, and potent heterologous rAd prime-boost regimens can be constructed using serologically distinct rAd vectors. We have recently demonstrated that a heterologous rAd26 prime/Ad5 boost regimen expressing SIV Gag afforded a durable 2.4 log reduction of setpoint viral loads after SIV<sub>MAC251</sub> challenge of Mamu-A\*01-negative rhesus monkeys, whereas a homologous rAd5 regimen provided no protection in this stringent challenge model<sup>90</sup>. These data suggest that vaccine candidates that elicit improved magnitude, breadth and quality of T lymphocyte responses may provide superior protective efficacy as compared with homologous rAd5 regimens.

### Perspectives and future directions

To a great extent, HIV-1 vaccine science is still in its infancy. Major unsolved problems remain, and a renewed commitment to basic discovery research in addition to preclinical studies and clinical trials will be required to move the field forward. Clinical trials that are focused on answering specific scientific hypotheses rather than exclusively aimed at product development may be most useful to the field at the present time. Certain vaccine regimens, such as heterologous

rAd prime-boost regimens, may offer the possibility of improved magnitude, breadth and quality of T lymphocyte responses as compared with the homologous rAd5 regimen. New antigen concepts, such as centralized consensus<sup>91,92</sup> and mosaic<sup>93</sup> immunogens, may also result in increased breadth of cellular immune responses and improved coverage of viral diversity.

Perhaps the most important research focus should be the development of improved Env immunogens to elicit broadly reactive neutralizing antibodies. Given the scope of this problem, increased basic research regarding the structure, function and immunogenicity of the Env glycoprotein will be required. Innovative and high-risk ideas should be pursued, and promising approaches should be tested as rapidly as possible in preclinical studies and eventually in clinical trials. Ultimately, it is likely that a combination vaccine consisting of separate vaccine components that elicit T lymphocytes and neutralizing antibodies will prove optimal. As a result, development of improved T-cell-based and antibody-based vaccine strategies should be pursued in parallel.

To achieve these goals, it will be critical to attract and to retain talented new investigators to the field. Funding programmes should therefore be expanded to encourage junior investigators to explore innovative ideas that address critical problems in the field. Given the scientific challenges currently facing the HIV-1 field, increased support and encouragement of fellows and junior faculty should be viewed as a top priority by both senior investigators and funding organizations. It will also be important for industry to continue to participate in the HIV-1 vaccine field, as biotechnology and pharmaceutical companies have critical knowledge and capacities that are not available in academia, government and non-profit organizations.

A current debate is whether the HIV-1 vaccine field can 'withstand' another vaccine efficacy study failure. For HIV-1, the scientific challenges are enormous, and thus so are the risks in testing any new vaccine concept. Clearly, the decision to advance a vaccine candidate into efficacy trials should be highly selective and based on a rigorous and transparent analysis of preclinical and clinical data. However, there is no way to determine whether a potentially promising vaccine candidate will afford protection in humans other than by conducting a clinical efficacy study. Multiple efficacy trials may be required, and many concepts will undoubtedly fail. We should therefore be ready to accept multiple failures of efficacy studies as part of the expected pathway towards the ultimate successful development of a safe and effective HIV-1 vaccine.

1. Barre-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**, 868–871 (1983).
2. Gallo, R. C. *et al.* Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224**, 500–503 (1984).
3. Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* **224**, 497–500 (1984).
4. Sarngadharan, M. G., Popovic, M., Bruch, L., Schupbach, J. & Gallo, R. C. Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS. *Science* **224**, 506–508 (1984).
5. Schupbach, J. *et al.* Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. *Science* **224**, 503–505 (1984).
6. Fauci, A. S. 25 years of HIV. *Nature* **453**, 289–290 (2008).
7. Quinn, T. C. *et al.* Viral load and heterosexual transmission of human immunodeficiency virus type 1. *N. Engl. J. Med.* **342**, 921–929 (2000).
8. Mascola, J. R. *et al.* Immunization with envelope subunit vaccine products elicits neutralizing antibodies against laboratory-adapted but not primary isolates of human immunodeficiency virus type 1. The National Institute of Allergy and Infectious Diseases AIDS Vaccine Evaluation Group. *J. Infect. Dis.* **173**, 340–348 (1996).
9. Moore, J. P. *et al.* Primary isolates of human immunodeficiency virus type 1 are relatively resistant to neutralization by monoclonal antibodies to gp120, and their neutralization is not predicted by studies with monomeric gp120. *J. Virol.* **69**, 101–109 (1995).
10. Flynn, N. M. *et al.* Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *J. Infect. Dis.* **191**, 654–665 (2005).

11. Pitisuttithum, P. *et al.* Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. *J. Infect. Dis.* **194**, 1661–1671 (2006).
12. Priddy, F. H. *et al.* Safety and immunogenicity of a replication-incompetent adenovirus type 5 HIV-1 clade B gag/pol/nef vaccine in healthy adults. *Clin. Infect. Dis.* **46**, 1769–1781 (2008).
13. Fauci, A. S. The release of new data from the HVTN 502 (STEP) HIV vaccine study. *NIH News* ([http://www3.niaid.nih.gov/about/directors/news/step\\_11707.htm](http://www3.niaid.nih.gov/about/directors/news/step_11707.htm)) (2007).
- These data demonstrate that a homologous rAd5-Gag/Pol/Nef vaccine regimen did not protect against HIV-1 in humans and may have increased risk of HIV-1 acquisition in individuals with pre-existing Ad5-specific neutralizing antibodies.**
14. Gaschen, B. *et al.* Diversity considerations in HIV-1 vaccine selection. *Science* **296**, 2354–2360 (2002).
15. Walker, B. D. & Korber, B. T. Immune control of HIV: the obstacles of HLA and viral diversity. *Nature Immunol.* **2**, 473–475 (2001).
16. Montefiori, D., Sattentau, Q., Flores, J., Esparza, J. & Mascola, J. Antibody-based HIV-1 vaccines: recent developments and future directions. *PLoS Med.* **4**, e348 (2007).
17. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648–659 (1998).
18. Wyatt, R. *et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**, 705–711 (1998).
19. Chen, B. *et al.* Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature* **433**, 834–841 (2005).
20. Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422**, 307–312 (2003).
21. Richman, D. D., Wrinn, T., Little, S. J. & Petropoulos, C. J. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl Acad. Sci. USA* **100**, 4144–4149 (2003).
22. Li, Y. *et al.* Broad HIV-1 neutralization mediated by CD4-binding site antibodies. *Nature Med.* **13**, 1032–1034 (2007).
23. Zhou, T. *et al.* Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445**, 732–737 (2007).
24. Haynes, B. F. *et al.* Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. *Science* **308**, 1906–1908 (2005).
25. Sun, Z. Y. *et al.* HIV-1 broadly neutralizing antibody extracts its epitope from a kinked gp41 ectodomain region on the viral membrane. *Immunity* **28**, 52–63 (2008).
26. Frey, G. *et al.* A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. *Proc. Natl Acad. Sci. USA* **105**, 3739–3744 (2008).
27. Baba, T. W. *et al.* Human neutralizing monoclonal antibodies of the IgG1 subtype protect against mucosal simian-human immunodeficiency virus infection. *Nature Med.* **6**, 200–206 (2000).
28. Mascola, J. R. *et al.* Protection of macaques against vaginal transmission of a pathogenic HIV-1/SIV chimeric virus by passive infusion of neutralizing antibodies. *Nature Med.* **6**, 207–210 (2000).
29. Pantaleo, G. *et al.* Major expansion of CD8<sup>+</sup> T cells with a predominant V beta usage during the primary immune response to HIV. *Nature* **370**, 463–467 (1994).
30. Koup, R. A. *et al.* Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J. Virol.* **68**, 4650–4655 (1994).
31. Borrow, P., Lewicki, H., Hahn, B. H., Shaw, G. M. & Oldstone, M. B. Virus-specific CD8<sup>+</sup> cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *J. Virol.* **68**, 6103–6110 (1994).
32. Musey, L. *et al.* Cytotoxic-T-cell responses, viral load, and disease progression in early human immunodeficiency virus type 1 infection. *N. Engl. J. Med.* **337**, 1267–1274 (1997).
33. Kiepiela, P. *et al.* Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* **432**, 769–775 (2004).
34. Kiepiela, P. *et al.* CD8<sup>+</sup> T-cell responses to different HIV proteins have discordant associations with viral load. *Nature Med.* **13**, 46–53 (2007).
35. Schmitz, J. E. *et al.* Control of viremia in simian immunodeficiency virus infection by CD8<sup>+</sup> lymphocytes. *Science* **283**, 857–860 (1999).
36. Jin, X. *et al.* Dramatic rise in plasma viremia after CD8<sup>+</sup> T cell depletion in simian immunodeficiency virus-infected macaques. *J. Exp. Med.* **189**, 991–998 (1999).
37. Phillips, R. E. *et al.* Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**, 453–459 (1991).
38. Allen, T. M. *et al.* Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. *Nature* **407**, 386–390 (2000).
39. Barouch, D. H. *et al.* Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. *Nature* **415**, 335–339 (2002).
40. Betts, M. R. *et al.* HIV nonprogressors preferentially maintain highly functional HIV-specific CD8<sup>+</sup> T cells. *Blood* **107**, 4781–4789 (2006).
41. Precopio, M. L. *et al.* Immunization with vaccinia virus induces polyfunctional and typically distinctive CD8<sup>+</sup> T cell responses. *J. Exp. Med.* **204**, 1405–1416 (2007).
42. Darrah, P. A. *et al.* Multifunctional TH1 cells define a correlate of vaccine-mediated protection against *Leishmania major*. *Nature Med.* **13**, 843–850 (2007).
43. Watkins, D. I., Burton, D. R., Kallas, E. G., Moore, J. P. & Koff, W. C. Nonhuman primate models and the failure of the Merck HIV-1 vaccine in humans. *Nature Med.* **14**, 617–621 (2008).
44. Seder, R. A., Darrah, P. A. & Roederer, M. T-cell quality in memory and protection: implications for vaccine design. *Nature Rev. Immunol.* **8**, 247–258 (2008).
45. Chun, T. W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**, 183–188 (1997).
46. Chun, T. W. *et al.* Early establishment of a pool of latently infected, resting CD4<sup>+</sup> T cells during primary HIV-1 infection. *Proc. Natl Acad. Sci. USA* **95**, 8869–8873 (1998).
47. Douek, D. C. *et al.* HIV preferentially infects HIV-specific CD4<sup>+</sup> T cells. *Nature* **417**, 95–98 (2002).
48. Veazey, R. S. *et al.* Gastrointestinal tract as a major site of CD4<sup>+</sup> T cell depletion and viral replication in SIV infection. *Science* **280**, 427–431 (1998).
49. Mattapallil, J. J. *et al.* Massive infection and loss of memory CD4<sup>+</sup> T cells in multiple tissues during acute SIV infection. *Nature* **434**, 1093–1097 (2005).
50. Li, Q. *et al.* Peak SIV replication in resting memory CD4<sup>+</sup> T cells depletes gut lamina propria CD4<sup>+</sup> T cells. *Nature* **434**, 1148–1152 (2005).
51. Brenchley, J. M. *et al.* Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nature Med.* **12**, 1365–1371 (2006).
52. Mattapallil, J. J. *et al.* Vaccination preserves CD4 memory T cells during acute simian immunodeficiency virus challenge. *J. Exp. Med.* **203**, 1533–1541 (2006).
53. Daniel, M. D., Kirchhoff, F., Czajak, S. C., Sehgal, P. K. & Desrosiers, R. C. Protective effects of a live attenuated SIV vaccine with a deletion in the *nef* gene. *Science* **258**, 1938–1941 (1992).
54. Wyand, M. S., Manson, K. H., Garcia-Moll, M., Montefiori, D. & Desrosiers, R. C. Vaccine protection by a triple deletion mutant of simian immunodeficiency virus. *J. Virol.* **70**, 3724–3733 (1996).
55. Learmont, J. C. *et al.* Immunologic and virologic status after 14 to 18 years of infection with an attenuated strain of HIV-1. A report from the Sydney Blood Bank Cohort. *N. Engl. J. Med.* **340**, 1715–1722 (1999).
56. Baba, T. W. *et al.* Pathogenicity of live, attenuated SIV after mucosal infection of neonatal macaques. *Science* **267**, 1820–1825 (1995).
57. Baba, T. W. *et al.* Live attenuated, multiply deleted simian immunodeficiency virus causes AIDS in infant and adult macaques. *Nature Med.* **5**, 194–203 (1999).
58. Murphey-Corb, M. *et al.* A formalin-inactivated whole SIV vaccine confers protection in macaques. *Science* **246**, 1293–1297 (1989).
59. Wille-Reece, U. *et al.* HIV Gag protein conjugated to a Toll-like receptor 7/8 agonist improves the magnitude and quality of Th1 and CD8<sup>+</sup> T cell responses in nonhuman primates. *Proc. Natl Acad. Sci. USA* **102**, 15190–15194 (2005).
60. Wille-Reece, U. *et al.* Toll-like receptor agonists influence the magnitude and quality of memory T cell responses after prime-boost immunization in nonhuman primates. *J. Exp. Med.* **203**, 1249–1258 (2006).
61. Casimiro, D. R. *et al.* Comparative immunogenicity in rhesus monkeys of DNA plasmid, recombinant vaccinia virus, and replication-defective adenovirus vectors expressing a human immunodeficiency virus type 1 *gag* gene. *J. Virol.* **77**, 6305–6313 (2003).
62. Graham, B. S. *et al.* Phase 1 safety and immunogenicity evaluation of a multiclade HIV-1 DNA candidate vaccine. *J. Infect. Dis.* **194**, 1650–1660 (2006).
63. Barouch, D. H. *et al.* Control of viremia and prevention of clinical AIDS in rhesus monkeys by cytokine-augmented DNA vaccination. *Science* **290**, 486–492 (2000).
64. Chong, S. Y. *et al.* Comparative ability of plasmid IL-12 and IL-15 to enhance cellular and humoral immune responses elicited by a SIVgag plasmid DNA vaccine and alter disease progression following SHIV(89.6P) challenge in rhesus macaques. *Vaccine* **25**, 4967–4982 (2007).
65. Luckay, A. *et al.* Effect of plasmid DNA vaccine design and *in vivo* electroporation on the resulting vaccine-specific immune responses in rhesus macaques. *J. Virol.* **81**, 5257–5269 (2007).
66. Liu, J., Kjekshus, R., Mathiesen, I. & Barouch, D. H. Recruitment of antigen-presenting cells to the site of inoculation and augmentation of human immunodeficiency virus type 1 DNA vaccine immunogenicity by *in vivo* electroporation. *J. Virol.* **82**, 5643–5649 (2008).
67. Shiver, J. W. *et al.* Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity. *Nature* **415**, 331–335 (2002).
68. Catanzaro, A. T. *et al.* Phase 1 safety and immunogenicity evaluation of a multiclade HIV-1 candidate vaccine delivered by a replication-defective recombinant adenovirus vector. *J. Infect. Dis.* **194**, 1638–1649 (2006).
69. Amara, R. R. *et al.* Control of a mucosal challenge and prevention of AIDS by a multiprotein DNA/MVA vaccine. *Science* **292**, 69–74 (2001).
70. Harari, A. *et al.* An HIV-1 clade C DNA prime, NYVAC boost vaccine regimen induces reliable, polyfunctional, and long-lasting T cell responses. *J. Exp. Med.* **205**, 63–77 (2008).
71. Shiver, J. W. & Emini, E. A. Recent advances in the development of HIV-1 vaccines using replication-incompetent adenovirus vectors. *Annu. Rev. Med.* **55**, 355–372 (2004).
72. Casimiro, D. R. *et al.* Attenuation of simian immunodeficiency virus SIVmac239 infection by prophylactic immunization with DNA and recombinant adenoviral vaccine vectors expressing Gag. *J. Virol.* **79**, 15547–15555 (2005).

This manuscript demonstrates that homologous rAd5 vaccine regimens were minimally effective against SIV<sub>MAC239</sub> challenges in rhesus monkeys.



73. Mothe, B. R. *et al.* Expression of the major histocompatibility complex class I molecule Mamu-A\*01 is associated with control of simian immunodeficiency virus SIVmac239 replication. *J. Virol.* **77**, 2736–2740 (2003).
74. Pal, R. *et al.* ALVAC-SIV-gag-pol-env-based vaccination and macaque major histocompatibility complex class I (A\*01) delay simian immunodeficiency virus SIVmac-induced immunodeficiency. *J. Virol.* **76**, 292–302 (2002).
75. Zhang, Z. Q. *et al.* Mamu-A\*01 allele-mediated attenuation of disease progression in simian-human immunodeficiency virus infection. *J. Virol.* **76**, 12845–12854 (2002).
76. Wilson, N. A. *et al.* Vaccine-induced cellular immune responses reduce plasma viral concentrations after repeated low-dose challenge with pathogenic simian immunodeficiency virus SIVmac239. *J. Virol.* **80**, 5875–5885 (2006).
77. Letvin, N. L. *et al.* Preserved CD4<sup>+</sup> central memory T cells and survival in vaccinated SIV-challenged monkeys. *Science* **312**, 1530–1533 (2006).
78. Vogels, R. *et al.* Replication-deficient human adenovirus type 35 vectors for gene transfer and vaccination: efficient human cell infection and bypass of preexisting adenovirus immunity. *J. Virol.* **77**, 8263–8271 (2003).
79. Abbink, P. *et al.* Comparative seroprevalence and immunogenicity of six rare serotype recombinant adenovirus vaccine vectors from subgroups B and D. *J. Virol.* **81**, 4654–4663 (2007).
80. Thorner, A. R. *et al.* Age dependence of adenovirus-specific neutralizing antibody titers in individuals from sub-Saharan Africa. *J. Clin. Microbiol.* **44**, 3781–3783 (2006).
81. Kostense, S. *et al.* Adenovirus types 5 and 35 seroprevalence in AIDS risk groups supports type 35 as a vaccine vector. *AIDS* **18**, 1213–1216 (2004).
82. Fauci, A. S. *et al.* HIV vaccine research: the way forward. *Science* **321**, 530–532 (2008).  
**This perspective describes revised NIH research priorities for HIV-1 vaccine research.**
83. Catanzaro, A. T. *et al.* Phase I clinical evaluation of a six-plasmid multiclade HIV-1 DNA candidate vaccine. *Vaccine* **25**, 4085–4092 (2007).
84. Fauci, A. S. NIAID will not move forward with the PAVE 100 HIV vaccine trial. *NIH News* (<http://www3.niaid.nih.gov/news/newsreleases/2008/pave100.htm>) (2008).
85. Barouch, D. H. *et al.* Immunogenicity of recombinant adenovirus serotype 35 vaccine in the presence of pre-existing anti-Ad5 immunity. *J. Immunol.* **172**, 6290–6297 (2004).
86. Roberts, D. M. *et al.* Hexon-chimaeric adenovirus serotype 5 vectors circumvent pre-existing anti-vector immunity. *Nature* **441**, 239–243 (2006).
87. Farina, S. F. *et al.* Replication-defective vector based on a chimpanzee adenovirus. *J. Virol.* **75**, 11603–11613 (2001).
88. Fitzgerald, J. C. *et al.* A simian replication-defective adenoviral recombinant vaccine to HIV-1 gag. *J. Immunol.* **170**, 1416–1422 (2003).
89. Liu, J. *et al.* Magnitude and phenotype of cellular immune responses elicited by recombinant adenovirus vectors and heterologous prime-boost regimens in rhesus monkeys. *J. Virol.* **82**, 4844–4852 (2008).
90. Barouch, D. H. Novel adenovirus vector-based vaccines for HIV-1. *Keystone Symposia on HIV Vaccines, Banff, Canada*. Abstract X7 009, page 60 (Keystone Symposia, 27 March–1 April 2008).
91. Liao, H. X. *et al.* A group M consensus envelope glycoprotein induces antibodies that neutralize subsets of subtype B and C HIV-1 primary viruses. *Virology* **353**, 268–282 (2006).
92. Weaver, E. A. *et al.* Cross-subtype T-cell immune responses induced by a human immunodeficiency virus type 1 group m consensus env immunogen. *J. Virol.* **80**, 6745–6756 (2006).
93. Fischer, W. *et al.* Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants. *Nature Med.* **13**, 100–106 (2007).  
**This manuscript proposes the use of polyvalent 'mosaic' antigens to improve immunologic coverage of global HIV-1 diversity.**

**Acknowledgements** The author would like to thank R. Dolin, N. Letvin, J. Mascola and J. McElrath for critically reviewing this manuscript. The author acknowledges support from the National Institutes of Health and the Bill & Melinda Gates Foundation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.H.B. ([dbarouch@bidmc.harvard.edu](mailto:dbarouch@bidmc.harvard.edu)).

# Photoemission kinks and phonons in cuprates

Arising from: F. Giustino, M. L. Cohen & S. G. Louie *Nature* **452**, 975–978 (2008)

One of the possible mechanisms of high transition temperature ( $T_c$ ) superconductivity is Cooper pairing with the help of bosons, which change the slope of the electronic dispersion as observed by photoemission. Giustino *et al.*<sup>1</sup> calculated that in the high temperature superconductor  $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$  crystal lattice vibrations (phonons) should have a negligible effect on photoemission spectra and concluded that phonons do not have an important role. Here we show that the calculations used by Giustino *et al.*<sup>1</sup> do not reproduce the huge influence of electron–phonon coupling on important phonons observed in experiments. Thus, we would similarly expect that these calculations do not explain the role of electron–phonon coupling for the electronic dispersion.

Density functional theory (DFT) calculations used by Giustino *et al.*<sup>1</sup> treat electrons and phonons as independent entities, which scatter each other. Because of this scattering, the electronic states acquire finite lifetimes and abrupt changes in dispersions (kinks) at the phonon energies. In addition, the phonons soften and broaden in energy. These effects are calculated from first principles without adjustable parameters. Therefore, if DFT is appropriate for the high  $T_c$  cuprates it ought to accurately reproduce the electronic contribution to phonon

softening and broadening deduced from neutron or X-ray scattering experiments.

DFT predicts that the phonon branch, in large part responsible for the calculated electronic dispersion kink, is the optical bond-stretching branch involving the bond-stretching motion of  $\text{CuO}_2$  plane oxygen against copper<sup>1</sup>. Several experimental papers have highlighted large anomalous renormalization of these phonons<sup>2–6</sup>. They have huge low temperature dispersion dips and/or line-width maxima around half-way ( $h = 0.3$ ) to the zone boundary in the superconductors  $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$  (refs 2 and 3) and  $\text{YBa}_2\text{Cu}_3\text{O}_7$  (ref. 4). However, DFT predicts a smooth dispersion without any pronounced features in either the dispersion or line width around  $h = 0.3$  (Fig. 1). Furthermore, the very small calculated line widths in Fig. 1b illustrate that the calculated electron–phonon coupling is very weak in absolute terms.

Substantial evidence points to an electronic origin of the phonon effect. First, the phonon anomaly weakens at elevated temperatures<sup>2,3</sup>, whereas alternatives such as phonon–phonon scattering and structural inhomogeneity should either show the opposite trend or have no temperature dependence. Second, the phonon effect appears at specific wavevectors and is phenomenologically similar to anomalies observed in conventional systems with strong electron–phonon coupling. Third, both phonon renormalization<sup>2,6</sup> and the photoemission kink<sup>7</sup> become bigger when hole concentration decreases from high doping (in which superconductivity is suppressed) towards so-called ‘optimal’ doping with the maximum superconducting  $T_c$ . This simultaneous enhancement of the two features may result from an increase in electron–phonon coupling due to enhanced electronic correlations or reduced screening not included in DFT. The findings of Giustino *et al.*<sup>1</sup> cannot rule out such hypotheses. The same holds for  $\text{YBa}_2\text{Cu}_3\text{O}_7$  where there is a similar disagreement between the experimental and DFT results for both the phonon dispersions and the photoemission kink<sup>8</sup>.

It is notable that many-body calculations predict a considerable enhancement of the coupling to bond-stretching phonons compared to DFT and describe anomalous doping dependence of the zone boundary phonons<sup>9,10</sup>, suggesting that strong correlation effects might be relevant. Recent high resolution photoemission measurements have found an oxygen isotope effect in the dispersion kink at the half-breathing phonon energy, hinting at an important role of oxygen phonons<sup>11</sup>. We conclude that more work is necessary to establish phonon contribution to the photoemission kink.

**D. Reznik<sup>1</sup>, G. Sangiovanni<sup>2</sup>, O. Gunnarsson<sup>3</sup> & T. P. Devereaux<sup>3</sup>**

<sup>1</sup>Forschungszentrum Karlsruhe, Institut für Festkörperphysik, PO Box 3640, D-76021 Karlsruhe, Germany.

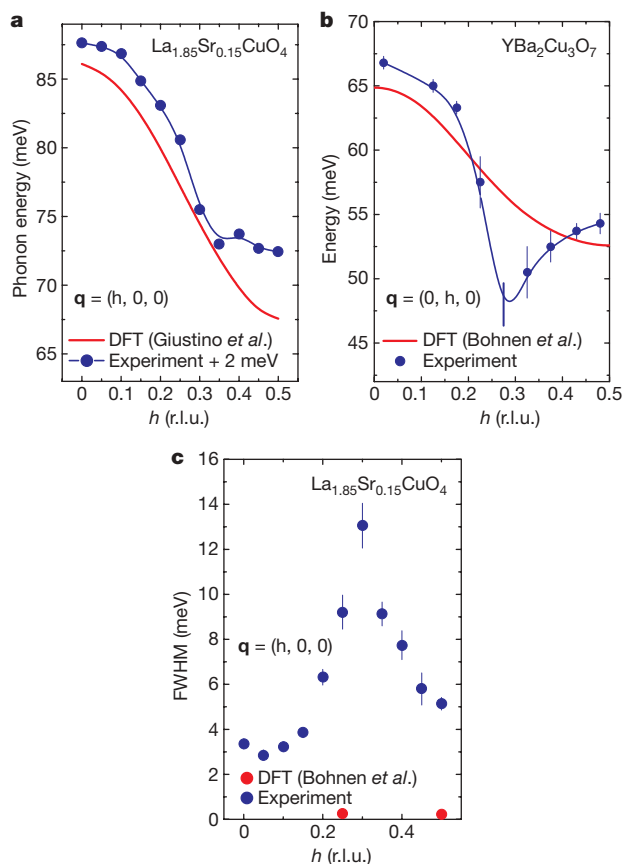
e-mail: reznik@llb.saclay.cea.fr

<sup>2</sup>Max-Planck-Institut für Festkörperforschung, D-70506 Stuttgart, Germany.

<sup>3</sup>Department of Photon Science, Stanford Linear Accelerator Center, Stanford University, 2575 Sand Hill Road, Menlo Park, California 94025, USA.

Received 1 July; accepted 20 August 2008.

- Giustino, F., Cohen, M. L. & Louie, S. G. Small phonon contribution to the photoemission kink in the copper oxide superconductors. *Nature* **452**, 975–978 (2008).
- Reznik, D. *et al.* Electron–phonon coupling reflecting dynamic charge inhomogeneity in copper oxide superconductors. *Nature* **440**, 1170–1173 (2006).
- Reznik, D. *et al.* Electron–phonon anomaly related to charge stripes: Static stripe phase versus optimally doped superconducting  $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$ . *J. Low Temp. Phys.* **147**, 353–364 (2007).
- Pintschovius, L. *et al.* Oxygen phonon branches in  $\text{YBa}_2\text{Cu}_3\text{O}_7$ . *Phys. Rev. B* **69**, 214506 (2004).
- Uchiyama, H. *et al.* Softening of Cu–O bond stretching phonons in tetragonal  $\text{HgBa}_2\text{CuO}_{4+\delta}$ . *Phys. Rev. Lett.* **92**, 197005 (2004).



**Figure 1 | Comparison of DFT predictions with experimental results for  $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$  and  $\text{YBa}_2\text{Cu}_3\text{O}_7$  at 10 K.** **a, b,** Experimental bond-stretching phonon dispersions<sup>2–4</sup> compared to DFT results<sup>1,12</sup>. The data in **a** are shifted by 2 meV. **c,** Phonon line widths in  $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$  (refs 2 and 3) compared with DFT results (K.-P. Bohnen, personal communication) on  $\text{YBa}_2\text{Cu}_3\text{O}_7$ . Giustino *et al.*<sup>1</sup> contains no line-width results for  $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$  but we expect them to be similar. Error bars represent s.d.;  $\mathbf{q}$  represents reduced wavevector in reciprocal lattice units (r.l.u.). FWHM, full-width at half-maximum.



6. Pintschovius, L., Reznik, D. & Yamada, K. Oxygen phonon branches in overdoped  $\text{La}_{1.7}\text{Sr}_{0.3}\text{Cu}_3\text{O}_4$ . *Phys. Rev. B* **74**, 174514 (2006).
7. Zhou, X. J. *et al.* Universal nodal Fermi velocity. *Nature* **423**, 398 (2003).
8. Heid, R., Bohnen, K.-P., Zeyher, R. & Manske, D. Momentum dependence of the electron-phonon coupling and self-energy effects in superconducting  $\text{YBa}_2\text{Cu}_3\text{O}_7$  within the local density approximation. *Phys. Rev. Lett.* **100**, 137001 (2008).
9. Rösch, O. & Gunnarsson, O. Electron-phonon interaction in the three-band model. *Phys. Rev. B* **70**, 224518 (2004).
10. Horsch, P. & Khaliullin, G. Doping dependence of density response and bond-stretching phonons in cuprates. *Physica B* **359–361**, 620–622 (2005).
11. Iwasawa, H. *et al.* An isotopic fingerprint of electron-phonon coupling in high- $T_c$  cuprates. Preprint at <<http://arxiv.org/abs/0808.1323>> (2008).
12. Bohnen, K.-P., Heid, R. & Krauss, M. Phonon dispersion and electron-phonon interaction for  $\text{YBa}_2\text{Cu}_3\text{O}_7$  from first-principles calculations. *Europhys. Lett.* **64**, 104–110 (2003).

doi:10.1038/nature07364

## ARTICLES

# Speciation through sensory drive in cichlid fish

Ole Seehausen<sup>1,2</sup>, Yohey Terai<sup>3</sup>, Isabel S. Magalhaes<sup>1,2</sup>, Karen L. Carleton<sup>4</sup>, Hillary D. J. Mrosso<sup>5</sup>, Ryutaro Miyagi<sup>3</sup>, Inke van der Sluijs<sup>6,†</sup>, Maria V. Schneider<sup>2,†</sup>, Martine E. Maan<sup>6,†</sup>, Hidenori Tachida<sup>7</sup>, Hiroo Imai<sup>8</sup> & Norihiro Okada<sup>3</sup>

Theoretically, divergent selection on sensory systems can cause speciation through sensory drive. However, empirical evidence is rare and incomplete. Here we demonstrate sensory drive speciation within island populations of cichlid fish. We identify the ecological and molecular basis of divergent evolution in the cichlid visual system, demonstrate associated divergence in male colouration and female preferences, and show subsequent differentiation at neutral loci, indicating reproductive isolation. Evidence is replicated in several pairs of sympatric populations and species. Variation in the slope of the environmental gradients explains variation in the progress towards speciation: speciation occurs on all but the steepest gradients. This is the most complete demonstration so far of speciation through sensory drive without geographical isolation. Our results also provide a mechanistic explanation for the collapse of cichlid fish species diversity during the anthropogenic eutrophication of Lake Victoria.

The sensory drive hypothesis for speciation<sup>1,2</sup> predicts that adaptation in sensory and signalling systems to different environments in allopatry may cause premating isolation on secondary contact of populations. Recent theoretical work suggested that sensory drive can lead to the evolution of colour polymorphisms<sup>3,4</sup> and speciation<sup>5</sup>, even in the absence of geographical isolation, when the light environment is heterogeneous. However, the only case of sympatric sister species, in which assortative mating has been shown to be facilitated by sensory drive, were sticklebacks in British Columbia<sup>6</sup>. Here we provide ecological, population genetic and molecular evidence for each of the predictions of sensory drive speciation<sup>2</sup> in sympatric cichlid fish inhabiting light gradients in Lake Victoria (East Africa).

Lake Victoria is spatially highly heterogeneous in water clarity and ambient light<sup>7,8</sup>, and there is much evidence that the cichlid visual system has been under strong diversifying selection during the adaptive radiation of cichlids into several hundred species in Lake Victoria<sup>9</sup>. Vertebrate visual pigments consist of a light-absorbing component, the chromophore, and a protein moiety, the opsin<sup>10</sup>. Spectral sensitivity is determined by the chromophore (A1 or A2 pigments), and by its interaction with the amino acid residues lining the retinal-binding pocket of the opsin in which the chromophore lies<sup>11</sup>. Eight different visual pigments have been found in all haplochromine cichlids<sup>12–14</sup>, but only a subset of these is expressed in any individual species<sup>12,14,15</sup>. Several *Pundamilia* species from Lake Victoria expressed the same complement of four opsin genes: short-wavelength-sensitive opsin gene 2a (SWS2A,  $\lambda_{\max}$  ~455 nm) in single cones; rhodopsin-like (RH2,  $\lambda_{\max}$  ~528 nm) and long-wavelength-sensitive opsin gene (LWS,  $\lambda_{\max}$  ~565 nm) in double cones; and rhodopsin (RH1,  $\lambda_{\max}$  ~505 nm) in rods<sup>16</sup>. Of these, the LWS opsin gene is by far the most variable among Lake Victoria cichlids<sup>13,17</sup>, with sequence variation being five times greater than

in Lake Malawi cichlids despite a tenfold greater age of the latter species flock<sup>18</sup>.

Female Lake Victoria cichlids have mating preferences for conspicuously coloured males<sup>19</sup>. Perception of conspicuousness is influenced by ambient and background light, signal transmission, receiver sensitivity and higher level processing<sup>2,20</sup>. Sympatric pairs of closely related cichlid species, one with red and one with blue nuptial colouration (Fig. 1 and Supplementary Fig. 3), are common in Lake Victoria<sup>8</sup>. Visual pigments have been compared for three pairs, and behavioural light detection thresholds measured in two. In each pair, the red species has its LWS  $\lambda_{\max}$  at a longer wavelength<sup>16,21</sup>, with a lower detection threshold for red but a higher one for blue light<sup>22,23</sup>. These observations are consistent with a role for sensory drive in speciation, whereby interaction between ambient light, natural-selection-driven divergence of visual sensitivities and sexual selection for conspicuous male colours leads to the fixation of different male colours<sup>1,2,16,23</sup>.

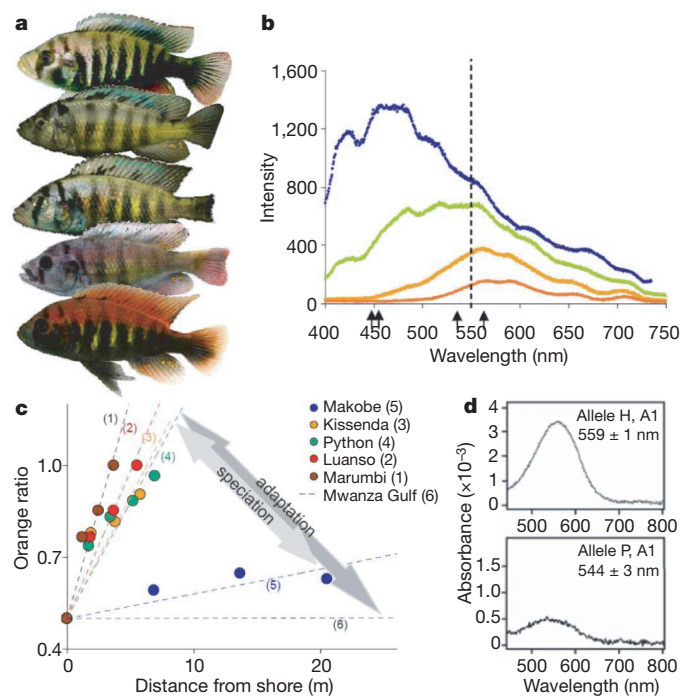
Examining the role of environmental gradients in speciation requires tests to replicate gradients, as is recognized both in evolutionary ecology<sup>24–26</sup> and in population genomics<sup>27</sup>. A recent model of clinal speciation through sensory drive<sup>5</sup>, as well as other models of clinal speciation<sup>28–30</sup>, predicts the greatest probability of speciation on gradients of intermediate slope. There, migration rates are sufficiently low to be compensated for by selection, but are sufficiently high to generate significant migration load<sup>31</sup> and intermediate genotypes with a poor fit to the local environment. Migration load and reduced fitness of intermediate genotypes lead to disruptive selection, which may be required for the evolution of assortative mating through reinforcement-like mechanisms<sup>28–30</sup>. Previously we demonstrated adaptive evolution in the LWS opsin gene of the Lake Victoria cichlid fish *Neochromis greenwoodi* and *Mbipia mbipi* along very shallow gradients of light colour mediated by variation in turbidity

<sup>1</sup>Institute of Zoology, University of Bern, Baltzerstr. 6, CH-3012 Bern, Switzerland. <sup>2</sup>Eawag, Swiss Federal Institute for Aquatic Science and Technology, Centre of Ecology, Evolution & Biogeochemistry, Department of Fish Ecology & Evolution, 6047 Kastanienbaum, Switzerland. <sup>3</sup>Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan. <sup>4</sup>Department of Biology, University of Maryland, College Park, Maryland 20742, USA. <sup>5</sup>Tanzania Fisheries Research Institute, Mwanza Centre, PO Box 475 Mwanza, Tanzania. <sup>6</sup>Department of Animal Ecology, Institute of Biology, Leiden University, PO Box 9516, 2300 RA Leiden, The Netherlands. <sup>7</sup>Department of Biology, Faculty of Sciences, Kyushu University, Ropponmatsu, Fukuoka 810-8560, Japan. <sup>8</sup>Department of Cellular and Molecular Biology, Primate Research Institute, Kyoto University, 484-8506 Japan. <sup>†</sup>Present addresses: Department of Biology, McGill University, 1205 Avenue Docteur Penfield, Montréal, Québec H3A 1B1, Canada (I.v.d.S.); The European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK (M.V.S.); University of Texas at Austin, Integrative Biology, 1 University Station C0930, Austin, Texas 78712, USA (M.E.M.).



between islands<sup>9</sup>. *LWS* genotype frequencies and male colour morph frequencies formed correlated clines, but, even though populations at opposite ends of one gradient fixed different *LWS* alleles, all populations retained polymorphism for colour, indicating that speciation remained incomplete<sup>9</sup>.

Here we investigate populations of cichlid fish living on light gradients primarily mediated by water depth within islands in Lake Victoria. *Pundamilia pundamilia* and *Pundamilia nyererei*<sup>22</sup> (Fig. 1a and Supplementary Fig. 3) are geographically fully sympatric. Within islands, they have narrowly parapatric depth ranges. Where they are phenotypically well differentiated, *P. pundamilia* has blue–grey male nuptial colouration whereas *P. nyererei* nuptial males are yellow with a bright crimson-red dorsum. Females of both are cryptically yellowish and have mating preferences for the nuptial colouration of conspecific males<sup>33,34</sup>. The red *P. nyererei* occurs at greater mean water depths, in more red-shifted ambient light than the blue *P. pundamilia*<sup>23</sup>. *P. nyererei* have a lower threshold for the detection of red light, whereas *P. pundamilia* possess a lower threshold for detection of blue light<sup>23</sup>. Earlier we found that red and blue fish tended to possess different alleles at the *LWS* opsin gene locus<sup>16</sup>. Here we fully develop this system to test predictions of sensory drive speciation.



**Figure 1 | Male phenotypes, light gradients and *LWS* opsin absorbance.** **a**, Variation in male nuptial colouration. Five phenotype classes from 0 ('blue', typical *P. pundamilia*; top) to 4 ('red', typical *P. nyererei*; bottom). **b**, An example of a moderately steep light gradient (Python island): surface light spectrum (blue) and three subsurface light spectra measured at 0.5 m (green), 1.5 m (orange) and 2.5 m (red) water depth. The line through 550 nm indicates the divide used to calculate the transmittance orange ratio. Arrows indicate peak absorbance of two opsin pigments: main allele groups at *LWS* opsin locus (544 nm and 559 nm) and known range of peak absorbance at *SWS2A* locus<sup>16</sup>. **c**, Slopes of seven different light gradients. The lines for two shallow gradients overlay each other and are together labelled 'Mwanza Gulf'. For this line, the x-axis represents the distance from clear water (rather than from shore). Significant differentiation in opsin genes was observed on all gradients with slopes equal to or shallower than the Kissenda (orange) line, but speciation was observed only on gradients with slopes between the Kissenda (orange) and the Makobe (blue) lines. The dark grey arrow indicates region with divergent adaption at *LWS* opsin gene, and the light grey arrow indicates region with speciation. **d**, Absorption spectra of *LWS* pigments evaluated by the dark–light difference spectra<sup>9</sup>. The *LWS* pigments were reconstituted from the H allele with A1 retinal (top) and from the P allele with A1 retinal (bottom).  $\lambda_{\max}$  values (with standard errors) are indicated.

If sensory drive caused speciation into a red and a blue species, we expected to find: (1) variation in the *LWS* opsin sequence at amino acid positions where they shift  $\lambda_{\max}$ ; (2) an association of such sequence variation with water depth, such that more red-shifted alleles occur at greater depth; and (3) an association of *LWS* alleles with the predominant male nuptial colouration of a population, such that populations with predominantly red-shifted opsin alleles have predominantly red males. Furthermore, if disruptive selection was required to complete speciation through the evolution of assortative mating, we predicted that the strongest associations between *LWS* alleles, water depth and colour occur on intermediate light slopes (prediction (4)). For testing prediction (4), we compared the data from the depth-mediated gradients of this study with data we had collected earlier on populations occupying the same depth at different islands with different turbidities<sup>9</sup> (see Supplementary Information).

### Light, depth and colour

We examined depth-mediated light gradients at five islands. The light climate of Lake Victoria is dominated by effects of particulate (non-phytoplankton) matter, selectively absorbing and scattering light of short wavelengths<sup>35</sup>, causing successive shifts of ambient light towards longer wavelengths with increasing water depth (this study), and also with increasing turbidity (earlier study)<sup>7,8</sup>. The rate at which ambient light changes with increasing depth is positively correlated with turbidity<sup>8</sup> (difference between islands in this study). The cichlids we study feed and breed right above and within the rocky substrate. We characterize depth-associated light gradients in their habitat by the change in the 'transmittance orange ratio' that occurs per metre as one moves outwards from the shore into the lake along the lake floor (the 'light slope', see Methods and Fig. 1b). Steeper slopes occur with more turbid water and steeper shores (Table 1). The steepest light slopes occurred at the most turbid sites, Marumbi and Luanso islands (Table 1 and Fig. 1c). Intermediate slopes occurred at Kissenda and Python islands, and the shallowest slope at Makobe island. The latter was still steeper than all the turbidity-mediated light slopes of our earlier work<sup>9</sup>. The size of the light differential between the ends of the gradients was similar between the five depth-mediated gradients, and larger than on the turbidity-mediated gradients (Table 1 and Supplementary Table 1).

Mapping the microdistribution of phenotypes on the five depth-mediated gradients using data from 960 males (Fig. 2a) revealed significant differences between islands. It showed the absence of any association between colour and ambient light (water depth) at Marumbi and Luanso (analysis of variance, ANOVA:  $df = 2$ ,  $F = 1.1$ ,  $P = 0.3$ , and  $df = 2$ ,  $F = 0.3$ ,  $P = 0.7$ , respectively), but significant associations at all other sites (ANOVA:  $df = 2$  (Kissenda),  $df = 1$  (Python, Makobe),  $F > 50$ ,  $P < 0.0001$ ), and increasing strength of association with decreasing light slope ( $F$  ratio against slope, logarithmic regression,  $df = 4$ ,  $R^2 = 0.87$ ,  $P = 0.021$ ; Fig. 3). Blue phenotypes are associated with shallow waters (<3 m) in all locations, whereas red phenotypes occur in shallow waters only on the steepest gradients, and become restricted to greater depths with decreasing light slope. Frequency distributions of male nuptial colour phenotypes differ significantly between islands too (Fig. 2b). Distributions are unimodal and skewed towards blue on the two steepest gradients. They are bimodal with few intermediates on gradients of intermediate slope, and consist of two discrete classes, blue and red, on the shallowest within-island gradient.

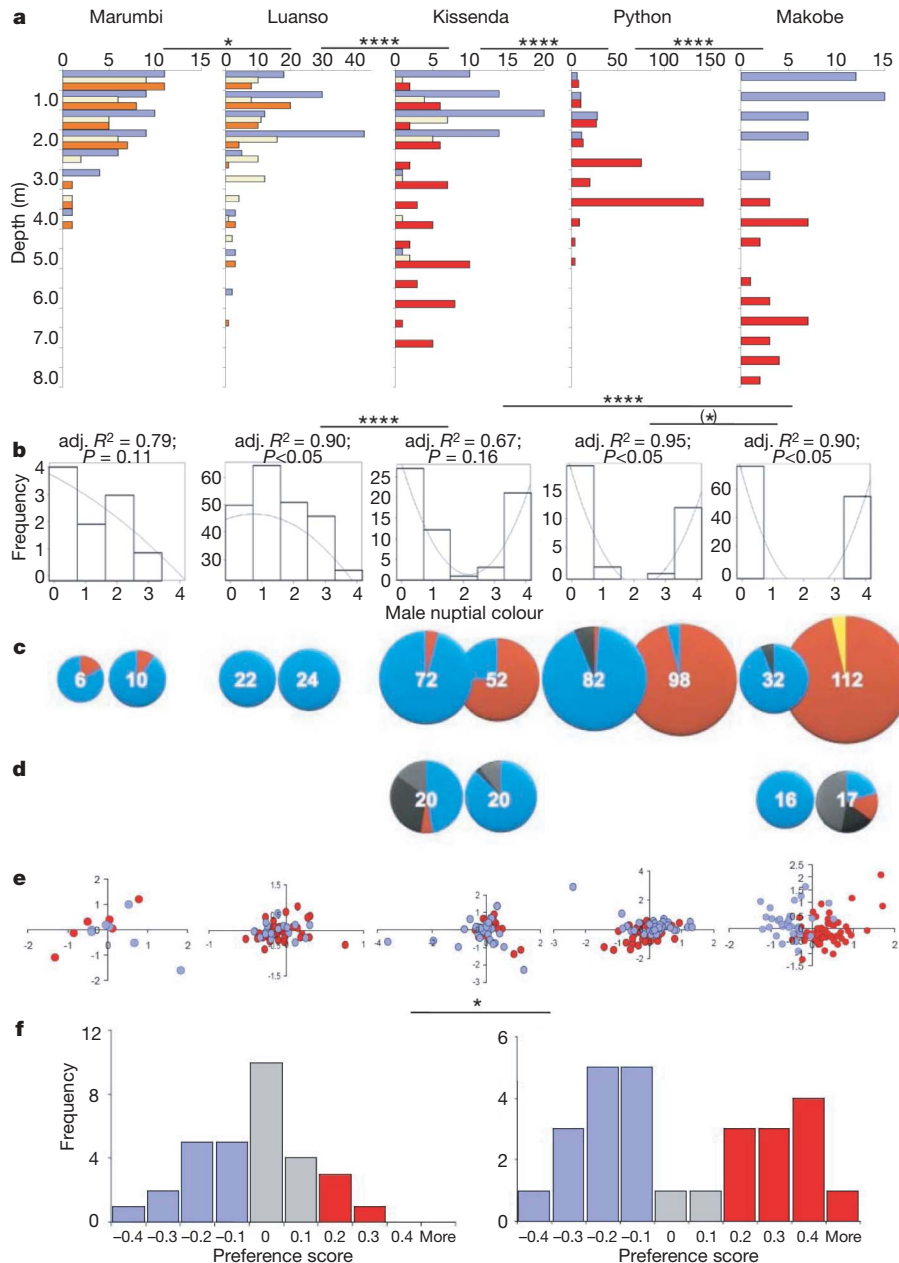
**Table 1 | The five environmental gradients of this study**

Island	Water clarity (cm Secchi) (mean $\pm$ s.d.)	Shoreline slope (mean $\pm$ s.d.)	Light slope	Light differential
Marumbi island	53 $\pm$ 8	0.82 $\pm$ 0.15	$1.4 \times 10^{-1}$	0.50
Luanso island	50 $\pm$ 10	0.54 $\pm$ 0.05	$9.6 \times 10^{-2}$	0.50
Kissenda island	78 $\pm$ 24	0.52 $\pm$ 0.12	$7.9 \times 10^{-2}$	0.50
Python island	96 $\pm$ 21	0.58 $\pm$ 0.24	$7.6 \times 10^{-2}$	0.50
Makobe island	225 $\pm$ 67	0.15 $\pm$ 0.04	$8 \times 10^{-3}$	0.35

**LWS gene variation, light and colour**

We observed 13 polymorphic sites (3 synonymous, 10 nonsynonymous) among the *LWS* sequences (Supplementary Table 6). Three nonsynonymous substitutions occurred at high frequencies. From

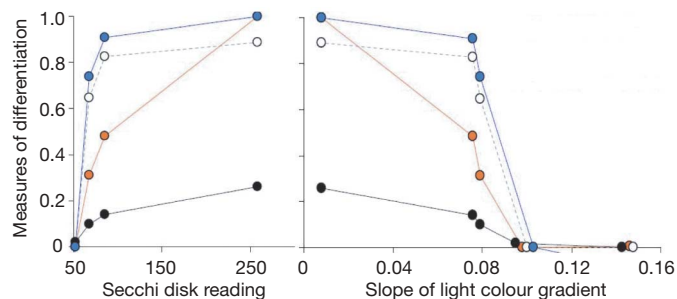
the bovine rhodopsin crystal structure<sup>36</sup> we inferred that two of these variable amino acid positions, 216 (nucleotide site 647) and 275 (823 and 824), are located in or near the retinal-binding pocket. The third one was position 230 (688), one of the tuning sites of human red/



**Figure 2 | Ecological, phenotypic, genetic and behavioural differentiation between blue and red *Pundamilia* nuptial phenotypes at five islands.** All data for the same island are presented in the same column. Significant differences between islands indicated by asterisks (all tests two-tailed): \* $P < 0.05$ , \*\*\*\* $P < 0.0001$ , (\*) $P < 0.1$ . **a**, Depth distributions of male nuptial colour phenotypes. Blue bars, blue; pale yellow bars, intermediate; and orange or red bars, red (orange if dominated by class 3; red if dominated by class 4). Significance levels of differences between islands in the divergence between red and blue reported as  $P$  values of  $G$ -tests. **b**, Frequency distributions of male nuptial colour phenotypes (see Fig. 1a and text). Lines are quadratic fits;  $R^2$  and significance levels indicated. Significance levels of differences between islands reported as  $P$  values of  $G$ -tests. **c**, Frequencies of functional allele groups at the *LWS* opsin gene by island and male colour (left, blue; right, red). Numbers report sample sizes of completely sequenced haplotypes. For Marumbi and Luanso islands, only the haplotypes of those individuals are included that could be assigned to 'blueish' and 'reddish' phenotypes (altogether 24 and 54 haplotypes were sequenced from Marumbi and Luanso,

respectively). Fish from Marumbi were divided into classes 0 + 1 and classes 2 + 3. Fish from Luanso were divided into classes 0 + 1 and 2–4. At all other islands, only fish of phenotype classes 0 and 4 were included. Alleles of the  $P$  group shown in blue, alleles of the  $H$  group in red, M3 alleles in yellow, and other alleles in grey. **d**, Allele frequencies at the *SWS2A* opsin gene by island and nuptial colour class. The *SWS2A*  $P$  allele shown in blue, the  $N$  allele in red, other alleles in black, and alleles not determined in grey. **e**, Individuals plotted on first and second axes of a factorial correspondence analysis of genetic variance calculated from 11 unlinked microsatellite loci. Colours indicate pooled male nuptial colour classes as described in **c**. **f**, Histograms of female mating preferences at Luanso island<sup>41</sup> (left) and Python island<sup>40</sup> (right, includes new data). Blue, preference classes in which most females had statistically significant individual preferences for blue males; red, preference classes in which most females had significant preferences for red males; grey, preference classes in which females had no significant mating preference. Significance level of the difference in the frequency distributions between the two islands reported as  $P$  value of a  $G$ -test.





**Figure 3 | Measures of differentiation between sympatric *Pundamilia* phenotypes plotted against water transparency (left) and light slope (right).** Blue symbols and line: Spearman rank correlations between colour and *LWS* genotype (best fit to water clarity  $R^2 = 0.79$ ,  $P = 0.045$ ,  $df = 4$ ; best fit to light slope  $R^2 = 0.69$ ,  $P(\text{one-tailed}) = 0.042$ ,  $df = 4$ ). Open symbols and dashed line: *LWS*  $F_{ST}$  between red and blue phenotypes (best fit to water clarity,  $R^2 = 0.79$ ,  $P = 0.044$ ,  $df = 4$ ; best fit to light slope,  $R^2 = 0.65$ ,  $P(\text{one-tailed}) = 0.045$ ,  $df = 4$ ). Filled orange symbols and orange line: association between colour and water depth (ANOVA  $F$  ratios (normalized to range 0–1 for display); best fit to water clarity  $R^2 = 0.99$ ,  $P = 0.000$ ; best fit to light slope  $R^2 = 0.87$ ,  $P = 0.021$ ; both  $df = 4$ ). Filled black symbols and black line: microsatellite  $F_{ST}$  (multiplied by 10 for display) between red and blue phenotypes (best fit to water clarity,  $R^2 = 0.99$ ,  $P = 0.000$ ; best fit to light slope,  $R^2 = 0.90$ ,  $P(\text{one-tailed}) = 0.013$ ; both  $df = 4$ ).

green opsin<sup>37</sup>. Focusing on these three positions, we divided alleles into three groups described previously<sup>9</sup>: the H group (all alleles with 216Y, 230A, 275C), the P group (216F, 230T, 275I) and the M3 group (216Y, 230T, 275I). M3 alleles can be considered recombinants or intermediate between H and P alleles. H and P alleles differed in only the 3 amino acid positions 216, 230 and 275. Substitutions at the other 7 nonsynonymous sites were rare and resulted in other allele variants (Supplementary Table 5).

We reconstituted the *LWS* pigments from P alleles *in vitro* with A1-derived retinal, and measured their absorption spectra, as previously done for the H alleles<sup>9</sup> (Fig. 1d). The peak spectral sensitivity ( $\lambda_{\text{max}}$ ) of the A1 pigment of the P allele was blue-shifted by 15 nm relative to the H allele. The  $\lambda_{\text{max}}$  values of cone outer segments expressing either P or H pigments were measured previously by microspectrophotometry, reporting too that P pigments were blue-shifted relative to H pigments<sup>16</sup>. Hence, the absorption spectra of P and H alleles seem to be adapted to shallower and deeper water light environments in Lake Victoria, respectively (Fig. 1b, d), supporting prediction (1).

Light gradients with slopes steeper than 0.09 were inhabited by populations with one or two different *LWS* alleles, whereas up to six different alleles were present on less steep gradients (Table 1 and Supplementary Tables 2 and 6). On these gradients of steepness  $0.008 \leq x \leq 0.09$ , H alleles were strongly associated with red nuptial

colouration ( $\chi^2 > 66$ ,  $df = 1$ ,  $P < 0.0001$ ; Spearman correlation coefficients 0.74, 0.91 and 1, respectively, for slopes 0.079, 0.076 and 0.008;  $P < 0.0001$ ), and were rare in blue phenotypes (Fig. 2c, Supplementary Table 6 and Supplementary Information), supporting prediction (3).

A strong association between *LWS* alleles and water depth emerges from these results, supporting prediction (2): at Marumbi and Luanso islands, most individuals reside in waters less than 3 m deep. P alleles strongly dominate. At all other islands, only the blue phenotype is confined to depths less than 3 m, and P alleles predominate among these fish, even where gene exchange with the red phenotype is frequent (see later). The sweep to high frequency of H alleles in the red phenotype is associated with shifting larger fractions of the population to depths beyond 3 m. At Kissenda island, 75% of the *LWS* alleles of the red population belong to the red-shifted H group. The proportion of H alleles, residing in red individuals, increases to Python island and further to Makobe island, associated with successively increasing fractions of the red population living in deep water (Fig. 2a versus 2c). Red and blue phenotypes were highly significantly differentiated at the *LWS* locus at Kissenda, Python and Makobe islands ( $F_{ST}$  (fixation index) 0.65, 0.83, 0.89), but neither at Luanso nor at Marumbi islands ( $F_{ST}$  0.00).

### Gene flow at neutral loci

A sensory drive model of speciation predicts that the rate of divergence at the opsin loci should exceed the rate of divergence at neutral loci. Our data are fully consistent with this prediction (Table 2). Pairwise  $F_{ST}$  between sympatric blue and red phenotypes estimated from 11 microsatellite loci reveal no differentiation at Marumbi or Luanso islands (Fig. 2e), consistent with the unimodal frequency distributions of male nuptial colour variants and the absence (Marumbi) or rarity of really red males. Pairwise  $F_{ST}$  at all other islands suggest significant, albeit weak, differentiation, consistent with the strongly bimodal frequency distributions of male nuptial colour variants and the emergence of the really red phenotype at those islands. Whereas  $F_{ST}$  at the *LWS* locus jumps from 0 at Marumbi and Luanso to 0.65 at Kissenda,  $F_{ST}$  at microsatellite loci increases gradually and much more slowly (Figs 2c, e and 3). The number of microsatellite loci carrying the signature of differentiation increases steadily from Marumbi and Luanso (0 out of 11) to Makobe island (7 out of 11; Table 2), consistent with the successive disappearance of intermediate phenotypes.

With the exception of Makobe island, all microsatellite  $F_{ST}$  among sympatric red and blue phenotypes are smaller than  $F_{ST}$  between any two allopatric populations of the blue phenotype, and 7 out of 10 of the red phenotype (Supplementary Fig. 1a and Supplementary Table 3). Even the largest between-phenotype  $F_{ST}$  at Makobe is smaller than most within-phenotype  $F_{ST}$  between islands. This suggests either more

**Table 2 | Pairwise  $F_{ST}$  statistics between sympatric phenotypes**

Island	Marumbi island	Luanso island	Kissenda island	Python island	Makobe island
Light slope	0.144	0.096	0.079	0.076	0.008
$F_{ST}$ at <i>LWS</i> opsin locus	0.000	0.000	<b>0.648</b>	<b>0.826</b>	<b>0.890</b>
$F_{ST}$ at microsatellite loci					
Ppun21	0.000	0.000	0.010	0.006	<b>0.023</b>
Ppun7	0.000	0.000	0.003	<b>0.023</b>	<b>0.013</b>
Ppun5	0.000	0.002	0.002	0.000	0.010
Ppun32	0.041	0.005	0.000	0.016	<b>0.080</b>
Ppun17	0.000	0.000	0.000	<b>0.046</b>	<b>0.027</b>
OSU16d	0.017	0.002	0.011	0.006	<b>0.020</b>
OSU20d	0.002	0.000	<b>0.040</b>	0.004	<b>0.008</b>
OSU19t	0.000	0.022	0.013	0.013	<b>0.032</b>
TMO5	0.000	0.013	0.000	0.012	0.010
Pzeb3	0.000	0.000	0.002	0.048	0.107
Pzeb5	0.000	0.000	0.024	0.024	0.049
Multilocus (11μsats)	0.000	0.002	<b>0.010</b>	<b>0.014</b>	<b>0.026</b>

Significant  $F_{ST}$  ( $P < 0.05$ ) are shown in bold.

gene flow or more recent divergence between phenotypes within islands than between island populations of the same phenotype. It implies either parallel maintenance of phenotypic differentiation in the face of gene flow, or parallel sympatric speciation. All H alleles as well as the most frequent P allele are shared with several distantly related cichlid species (Supplementary Fig. 4). The two *Pundamilia* H alleles are the most frequent H alleles in those distantly related species too. Either red *Pundamilia* populations acquired these alleles once or multiple times from other species through introgressive hybridization, or the shared ancestor of red and blue *Pundamilia* possessed all the P and H alleles. In either scenario, the H and P allele split must pre-date the origin of the blue and red *Pundamilia* species.

### Selection on the LWS gene

We analysed sequences up- and down-stream of *LWS* in a population (Python) that exhibits strong divergence in *LWS* but only weak differentiation at microsatellite loci. Sliding-window  $F_{ST}$  analysis revealed at least 6 times greater divergence in the *LWS* gene exons and in 2 kilobases (kb) of upstream sequence ( $F_{ST} > 0.8$ ; Supplementary Fig. 2a) than in the downstream sequences ( $F_{ST} < 0.15$ ), and more than 50 times greater divergence than at microsatellite loci (Table 2). Together with results of McDonald tests<sup>38</sup> and HKA tests<sup>39</sup> (Supplementary Table 4 and Supplementary Information), this is consistent with a recent selective sweep in the red species, associated with increased presence in a red-shifted environment.

### Divergence in the SWS2A opsin gene

We sequenced the SWS2A opsin gene at two islands to test for divergence at the short-wavelength end of the light spectrum. Out of 10 variable nucleotide positions, 5 were synonymous and 5 were located in introns (Supplementary Table 7). At Kissenda, the SWS2A sequences were variable in both phenotypes, and differentiated between them ( $F_{ST} 0.1$ ,  $P < 0.01$ ). At Makobe, a single SWS2A sequence variant was almost fixed in *P. pundamilia*, and the species were more strongly differentiated, although not as strongly as in *LWS* ( $F_{ST}$ : 0.437,  $P < 0.001$ ; Fig. 2d).

### Female mating preferences

Experiments and field data suggest that female *Pundamilia* use male colour as an important mate choice cue<sup>19,33,34</sup>. Most wild and laboratory-bred Python island females prefer either blue or red males, but laboratory-bred  $F_1$ -hybrid females, most laboratory-bred  $F_2$ -hybrid females and most Luanso females have no preference between blue and red males<sup>40,41</sup>. Combining published data<sup>40,41</sup> with previously unpublished data for 11 females from Python island, we find that the frequency distributions of female mating preferences differ between the islands ( $G$ -test,  $P = 0.02$ ), roughly resembling those of male nuptial colour (compare Fig. 2f with 2b). The distribution at Luanso (38 females) had a single mode on no preference, and a skew towards blue preference. The distribution at Python (27 females) was bimodal.

We analysed Python island non-hybrid and laboratory-bred  $F_2$  hybrid females to ask whether the *LWS* genotype directly determines mating preference. For non-hybrids and hybrids combined, we observed a significant association between individual *LWS* genotype and mating preference ( $\chi^2 = 22$ ,  $df = 10$ ,  $P = 0.03$ , 10,000 randomizations). However, this relationship was not significant when restricted to  $F_2$  hybrid females ( $\chi^2 = 10.2$ ,  $df = 6$ ,  $P = 0.13$ , 10,000 randomizations). Hence, variation in the SWS2A-SWS2B-*LWS* chromosomal region alone does not strongly predict visual mating preferences in a laboratory environment: some component of mating preference seems independent of it, consistent with biometric estimates that implied that the difference in mating preferences between *P. pundamilia* and *P. nyererei* was due to more than one factor<sup>40</sup>. Modelling light detection, using solar spectrum, water transmission, *Pundamilia* colour patch reflection and *Pundamilia* visual pigment absorption, suggested that a  $\lambda_{max}$  shift of 4 nm towards longer

wavelengths causes a 10% increase in quantum catch for a fish looking at a red patch<sup>16</sup>. It seems probable that, in interaction with ambient light in the natural environment, the opsin genotype more strongly determines mating preference than it does under standard laboratory light conditions.

### Discussion

Our data on ambient light colour, male nuptial colour, visual pigment  $\lambda_{max}$  and female mating preference indicate sensory drive speciation, which occurred or is maintained by selection without geographical isolation. However, we only observed this under a restricted range of environmental conditions. At all sites with moderately shallow to moderately steep light gradients, two differentiated populations emerged with strong associations between water depth, *LWS* alleles, colouration and preferences (Fig. 3). Strong bimodalities in the quantitative traits colour and preference, strong heterozygote deficiencies at the *LWS* opsin gene, and differentiation at microsatellite loci clearly indicate speciation initiated by strong selection on *LWS*. Very steep light gradients, in contrast, were inhabited by single panmictic populations that showed little variation in *LWS*, even though they contained some variation in colour and mating preference.

The following sensory drive speciation scenario is fully consistent with our data. First, divergent natural selection between light regimes at different water depths acts on *LWS*. Second, sexual selection for conspicuous colouration is also divergent because perceptual biases differ between light regimes. Third, their interaction generates initial deviation from linkage equilibrium between *LWS* and nuptial colour alleles as observed on all but the steepest gradients. Fourth, subsequent disruptive selection due to reduced fitness of genotypes with a mismatch between *LWS* and colour alleles causes speciation, perhaps involving reinforcement-like selection for mating preferences, whereby male nuptial colour may serve as a marker trait for opsin genotype.

The strong association between *LWS* alleles and male nuptial colouration with few or no mismatch genotypes in sympatric species pairs is not restricted to *P. pundamilia* and *P. nyererei* (Table 3, Supplementary Fig. 3 and Supplementary Information). In contrast with these results, we did not find any such discontinuities in the frequency distribution of opsin genotypes along very shallow (between-island) gradients investigated earlier<sup>9</sup>—that is, intermediate *LWS* genotypes predominated in large sections of each gradient. This suggested the presence of divergent selection but the absence of disruptive selection (or the absence of an evolutionary response to disruptive selection). This is consistent with the low migration load predicted from the very small difference in ambient light that migrants between adjacent islands experience (Supplementary Table 1). Despite positive correlations between frequencies of *LWS* alleles and male nuptial colour morphs, and complete fixation of different *LWS* alleles between some populations, speciation as would be indicated first, by strong association between *LWS* and colour and, second, by genotypic and phenotypic discontinuities was not observed on these gradients. This may be due to a difference between the taxa that we studied, but it may also imply that speciation requires disruptive selection, and hence migration and gene flow between habitats<sup>5,28–30,42</sup>. In contrast, when migration exceeds selection, divergence cannot occur either<sup>43,44</sup>. This explains the absence of speciation on the steepest of our gradients.

Our results are relevant to conservation because they provide a mechanistic explanation for the collapse of cichlid fish species diversity during the anthropogenic eutrophication of Lake Victoria<sup>8</sup>. Eutrophication changes the slope of environmental light gradients, and, by steepening them, potentially moves sites from the region in parameter space that is permissive of species coexistence into the region that is not. We hope these results help focus attention of biodiversity conservation efforts in Lake Victoria and other lakes to issues of water quality.



**Table 3 | LWS opsin allele-group frequency (%) and male nuptial colouration in species of *Pundamilia***

Species	Population	Male nuptial colour type	P	M3	H	Others	n†
<i>P. "Luanso"</i>	Luanso island	Predominantly blue	100	0	0	0	54
<i>P. "Marumbi"</i>	Marumbi island	Predominantly blue	92	0	8	0	24
<i>P. pundamilia</i>	Makobe island	Blue	94	0	0	6	32
<i>P. pundamilia</i>	Igombe island	Blue	83	17	0	0	6
<i>P. pundamilia</i> -like*	Kissenda island	Blue	96	0	4	0	70
<i>P. pundamilia</i> -like*	Python island	Blue	90	1	4	5	82
<i>P. azurea</i> <sup>16</sup>	Ruti island	Blue	100	0	0	0	6
<i>P. nyererei</i> -like*	Kissenda island	Red dorsum	25	0	75	0	52
<i>P. nyererei</i> -like*	Python island	Red dorsum	4	0	96	0	98
<i>P. nyererei</i>	Makobe island	Red dorsum	0	4	96	0	112
<i>P. igneopinnis</i>	Igombe island	Red dorsum	0	0	100	0	6
<i>P. "red head"</i> <sup>16</sup>	Zue island	Red chest	0	0	100	0	6
Total							548

\* Hybridizing populations (neither *P. pundamilia* nor *P. nyererei*, but the hybridizing blue (*P. pundamilia*-like) and red (*P. nyererei*-like) populations shown in Fig. 2 (this study)).

† n represents n haplotypes sequenced.

## METHODS SUMMARY

Ambient, absorbance and transmittance light spectra were measured with an Ocean Optics PS 1000 spectrophotometer and a 100 µm optical fibre, in the shade between 8:50 and 9:00 in the morning. We calculated the 'transmittance orange ratio' as the ratio of transmittance in the 550–700 nm range over the total visible range. The 'light slope' was obtained by regressing the transmittance orange ratio against distance (m) from the shore along the lake floor. Male fish in breeding colouration were collected by angling and netting; 480 males were photographed immediately in a photo cuvette. Water depth was measured and recorded to the nearest 0.5 m for each of 960 males. To determine the functional relevance of the observed amino acid substitutions in the LWS genes, the sequence of the P allele was reconstructed from the H allele by *in vitro* mutagenesis. The pigments were then expressed, reconstituted and purified as described elsewhere<sup>9</sup>. Absorption spectra of reconstituted pigments were measured before and after irradiation with light (>490 nm). DNA was extracted from fin tissue of 305 individuals and amplified using 11 microsatellite primers. The fragments were analysed on a Beckman Coulter CEQ 8000 Genetic Analysis System. Determination of the opsin genes was as described previously<sup>13</sup>. We sequenced exons 2–5 of LWS (872 bp), which encode the trans-membrane region, from 263 individuals (526 haplotypes). We sequenced exons 1–5 (including introns) of the SWS2A gene from males of Makobe island and Kissenda island. For detection of selection, the LWS gene and its 5-kb upstream and 3.5-kb downstream flanking sequences (total 10.5 kb) were amplified by long PCR<sup>9</sup> and sequenced from 10 red and 9 blue males from Python island. To determine female mating preferences, we conducted laboratory two-way mate choice assays with females from Luanso island and Python island and laboratory-bred F<sub>2</sub> hybrids from Python island<sup>40</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 24 April; accepted 25 July 2008.**

- Schluter, D. & Price, T. Honesty, perception and population divergence in sexually selected traits. *Proc. R. Soc. Lond. B* **253**, 117–122 (1993).
- Boughman, J. W. How sensory drive can promote speciation. *Trends Ecol. Evol.* **17**, 571–577 (2002).
- Gray, S. M. & McKinnon, J. S. Linking color polymorphism maintenance and speciation. *Trends Ecol. Evol.* **22**, 71–79 (2007).
- Chunco, A. J., McKinnon, J. S. & Servedio, M. R. Microhabitat variation and sexual selection can maintain male color polymorphisms. *Evolution* **61**, 2504–2515 (2007).
- Kawata, M., Shoji, A., Kawamura, S. & Seehausen, O. A genetically explicit model of speciation by sensory drive within a continuous population in aquatic environments. *BMC Evol. Biol.* **7**, 99 (2007).
- Boughman, J. W. Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature* **411**, 944–948 (2001).
- Levring, T. & Fish, G. R. The penetration of light in some tropical East African waters. *Oikos* **7**, 98–109 (1956).
- Seehausen, O., van Alphen, J. J. M. & Witte, F. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808–1811 (1997).
- Tera, Y. *et al.* Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biol.* **4**, 2244–2251 (2006).
- Shichida, Y. *The Retinal Basis of Vision: Visual pigment: photochemistry and molecular evolution* (ed. Toyoda, J.-I.) 23–35 (Elsevier Science, 1999).
- Yokoyama, S., Blow, N. S. & Radlwiner, F. B. Molecular evolution of color vision of zebra finch. *Gene* **259**, 17–24 (2000).
- Carleton, K. L. & Kocher, T. D. Cone opsin genes of African cichlid fishes: Tuning spectral sensitivity by differential gene expression. *Mol. Biol. Evol.* **18**, 1540–1550 (2001).

- Tera, Y., Mayer, W. E., Klein, J., Tichy, H. & Okada, N. The effect of selection on a long wavelength-sensitive (LWS) opsin gene of Lake Victoria cichlid fishes. *Proc. Natl Acad. Sci. USA* **99**, 15501–15506 (2002).
- Parry, J. W. L. *et al.* Mix and match color vision: Tuning spectral sensitivity by differential opsin gene expression in Lake Malawi cichlids. *Curr. Biol.* **15**, 1734–1739 (2005).
- Carleton, K. *et al.* Visual sensitivities tuned by heterochronic shifts in opsin gene expression. *BMC Biol.* **6**, 22 (2008).
- Carleton, K. L., Parry, J. W. L., Bowmaker, J. K., Hunt, D. M. & Seehausen, O. Colour vision and speciation in Lake Victoria cichlids of the genus *Pundamilia*. *Mol. Ecol.* **14**, 4341–4353 (2005).
- Spady, T. C. *et al.* Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Mol. Biol. Evol.* **22**, 1412–1422 (2005).
- Genner, M. J. *et al.* Age of cichlids: New dates for ancient lake fish radiations. *Mol. Biol. Evol.* **24**, 1269–1282 (2007).
- Maan, M. E. *et al.* Intraspecific sexual selection on a speciation trait, male coloration, in the Lake Victoria cichlid *Pundamilia nyererei*. *Proc. R. Soc. Lond. B* **271**, 2445–2452 (2004).
- Endler, J. A. Some general comments on the evolution and design of animal communication systems. *Phil. Trans. R. Soc. Lond. B* **340**, 215–225 (1993).
- Vandermeer, H. J., Anker, G. C. & Barel, C. D. N. Ecomorphology of retinal structures in zooplanktivorous haplochromine cichlids (Pisces) from Lake Victoria. *Environ. Biol. Fishes* **44**, 115–132 (1995).
- Smit, S. A. & Anker, G. C. Photopic sensitivity to red and blue light related to retinal differences in two zooplanktivorous haplochromine species (Teleostei, Cichlidae). *Neth. J. Zool.* **47**, 9–20 (1997).
- Maan, M. E., Hofker, K. D., van Alphen, J. J. M. & Seehausen, O. Sensory drive in cichlid speciation. *Am. Nat.* **167**, 947–954 (2006).
- Endler, J. A. Gene flow and population differentiation. *Science* **179**, 243–250 (1973).
- Schluter, D. & Nagel, L. M. Parallel speciation by natural selection. *Am. Nat.* **146**, 292–301 (1995).
- Nosil, P., Egan, S. R. & Funk, D. J. Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**, 316–336 (2008).
- Stinchcombe, J. T. & Hoekstra, H. E. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**, 158–170 (2008).
- Doebeli, M. & Dieckmann, U. Speciation along environmental gradients. *Nature* **421**, 259–264 (2003).
- Gavrilets, S. *Fitness Landscapes and the Origin of Species* (Princeton Univ. Press (2004)).
- Leimar, O., Doebeli, M. & Dieckmann, U. Evolution of phenotypic clusters through competition and local adaptation along an environmental gradient. *Evolution* **62**, 807–822 (2008).
- Nosil, P., Vines, T. H. & Funk, D. J. Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**, 705–719 (2005).
- Seehausen, O. *Lake Victoria Rock Cichlids. Taxonomy, Ecology and Distribution*. (Verduijn Cichlids, 1996).
- Seehausen, O. & van Alphen, J. J. M. The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (*Haplochromis nyererei* complex). *Behav. Ecol. Sociobiol.* **42**, 1–8 (1998).
- Stelkens, R. B., Pierotti, M. E. R., Joyce, D. A., Smith, A. M., van der Sluijs, I. & Seehausen, O. Female mating preferences facilitate disruptive sexual selection on male nuptial colouration in hybrid cichlid fish. *Phil. Trans. R. Soc. B* **363**, 2861–2870 (2008).
- Okullo, W. *et al.* Parameterization of the inherent optical properties of Murchison Bay, Lake Victoria. *Appl. Opt.* **46**, 8553–8561 (2007).
- Palczewski, K. *et al.* Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739–745 (2000).
- Asenjo, A. B., Rim, J. & Oprian, D. D. Molecular determinants of human red/green color discrimination. *Neuron* **12**, 1131–1138 (1994).

38. McDonald, J. H. Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol. Biol. Evol.* **15**, 377–384 (1998).
39. Hudson, R. R., Kreitman, M. & Aguade, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
40. Haesler, M. P. & Seehausen, O. Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: implications for speciation. *Proc. R. Soc. B* **272**, 237–245 (2005).
41. van der Sluijs, I., van Alphen, J. J. M. & Seehausen, O. Preference polymorphism for coloration but no speciation in a population of Lake Victoria cichlids. *Behav. Ecol.* **19**, 177–183 (2008).
42. Nosil, P., Crespi, B. J. & Sandoval, C. P. Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. *Proc. R. Soc. Lond. B* **270**, 1911–1918 (2008).
43. Nosil, P. & Crespi, B. J. Does gene flow constrain adaptive divergence or vice versa? A test using ecomorphology and sexual isolation in *Timema cristinae* walking-sticks. *Evolution* **58**, 102–112 (2004).
44. Rasanen, K. & Hendry, A. Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecol. Lett.* **11**, 624–636 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge the Tanzania Commission for Science & Technology for research permissions, the Tanzania Fisheries Research Institute, and its Muranza Centre director E. F. B. Katunzi, for hospitality and logistical support; M. Kayeba, M. Haluna, S. Mwaiko, M. Haesler and E. Burgerhout for help

with data and fish collection; H. Araki, L. Excoffier, L. Harmon, B. Ibelings, I. Keller, T. Kocher, P. Nosil, M. Pierotti, D. Schluter, A. Sivasundar and O. Svensson for comments on the manuscript; and M. Kawata, J. J. M. van Alphen, K. Young, R. Stelkens and E. Bezault for discussion. This work was supported by Swiss National Science Foundation project 3100A0-106573 (to O.S.), and by the Ministry of Education, Culture, Sports, Science and Technology of Japan (to N.O.).

**Author Contributions** O.S. conceived and designed the study, collected, photographed and identified fish, measured light and shore slopes, supervised field work, conducted the hybridization experiments, supervised microsatellite analyses and mate choice experiments, and did the statistical data analyses and the writing. Y.T. designed experiments on opsins, did most of the laboratory work and data analysis on opsins, and contributed to writing. I.S.M. collected depth distribution data and did all microsatellite analyses. K.L.C. determined *LWS* sequences from experimental females and contributed to writing. H.D.J.M. collected depth distribution, light data and fish. R.M. determined *LWS* and *SWS2A* sequences with Y.T. I.v.d.S. collected fish and conducted mate choice experiments. M.V.S. helped with the microsatellite analysis. M.E.M. collected fish and measured light. H.T. performed analysis of selection pressure with Y.T. H.I. measured opsin pigment absorbance with Y.T. N.O. designed and supervised the laboratory work on opsins and contributed to the writing.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to O.S. ([ole.seehausen@aqua.unibe.ch](mailto:ole.seehausen@aqua.unibe.ch)) or N.O. ([nokada@bio.titech.ac.jp](mailto:nokada@bio.titech.ac.jp)).

## METHODS

**Ambient light gradients and water clarity.** Water transparency was measured using a white Secchi disk. Ambient, absorbance and transmittance light spectra between 400 nm and 750 nm were measured every metre between the surface and 3 m water depth with an Ocean Optics PS 1000 spectrophotometer and an optical fibre (100 µm), using SpectraWin 4.16 software (Avantes). Measurements were taken in the shade, between 8:50 and 9:00 in the morning. We calculated at every depth the 'transmittance orange ratio', which is a property of the water unaffected by variation in solar irradiance, as the ratio of transmittance in the 550–700 nm range (yellow, orange, red) over the total visible range (400–700 nm). The steepness of the light gradient, the 'light slope', was calculated by regressing the transmittance orange ratio against the mean distance (m) from the shore, measured along the lake floor in three transects for every island. The turbidity-mediated between-island light slopes were calculated by regressing the transmittance orange ratio measured at every island at 2 m water depth against the distance (m) from the clear water end of each gradient. The light differential was measured for both types of gradients as the difference between the transmittance orange ratios at the end points of a gradient. The largest possible value is 0.5, which is given when there is no longer any detectable blue light at the deep end of a gradient (transmittance orange ratio = 1 (that is, orange is the only transmitted light); whereas at the surface the full amounts of both blue and orange light are present (that is, transmittance orange ratio = 0.5)).

**Frequency and depth distribution of male colouration.** Males were collected by angling and gill nets in April and August 2001, February 2003, and January and May 2005. Photos were taken of 11 (Marumbi), 241 (Luanso), 64 (Kissenda), 34 (Python) and 130 (Makobe) males in breeding dress—480 in total—immediately on capture in specially designed photographic cuvettes. Photos were scored on a 5-point (0–4) colour phenotype scale by two to five independent observers, and the mean value was used<sup>41</sup> (Fig. 1). Phenotype scoring of different observers was very similar (Spearman correlations between 0.605 and 0.729,  $P < 0.05$ ). Linear regressions with a quadratic term were fitted to the log-transformed counts of the colour phenotypes from each island separately using  $R^{45}$ . Frequency distributions were compared between islands by  $G$ -tests.

Water depth was measured and recorded to the nearest 0.5 m for each of 960 males. The association between phenotype and water depth was tested for each island separately using ANOVA tests. These males were assigned to colour classes in the field, and only three robust classes were used: blue, intermediate and red (corresponding to classes 0 + 1, 2 and 3 + 4).  $G$ -tests were performed to compare depth distributions between islands. The curve-fitting procedure in SPSS (SPSS Inc. 2005) was used to quantify the relationship between strength of association ( $F$ -value) and steepness of the light slope.

**LWS absorption spectra.** *In vitro* mutagenesis of *LWS* for construction of the sequence of P alleles, expression, reconstitution, purification and measurement were performed as described previously<sup>9</sup> with minor modifications. We measured absorption spectra of reconstituted pigments before and after irradiation with light ( $>490$  nm). On the basis of the  $\lambda_{\max}$  values determined by 3 independent difference spectra calculated from the measurements using independent preparations, we determined the absorption maximum values for each allele with standard errors.

**Population genetics of neutral loci.** DNA was extracted from fin tissue of 305 individuals (Marumbi 13, Luanso 61, Kissenda 59, Python 84, Makobe 88) and amplified using 11 microsatellite primers developed for these or other

haplochromine species (see Supplementary Methods). We used Arlequin<sup>46</sup> to calculate observed and expected heterozygosities, to test for significance of departure from Hardy–Weinberg equilibrium for each locus in each population (1 million MCMC permutations), and for significant deviations from linkage equilibrium (10,000 permutations). After sequential Bonferroni correction<sup>47</sup>, 3 out of 55 tests revealed significant deviations from Hardy–Weinberg equilibrium (1 locus each in *P. pundamilia* and *P. nyererei* from Makobe, 1 in *P. pundamilia* from Kissenda), and 2 tests of linkage equilibrium were significant: 1 in *P. pundamilia* from Python island and 1 in *P. pundamilia* from Kissenda island. Because there was no indication of any consistent linkage disequilibrium across populations between any pair of loci, all loci were retained for subsequent analysis. Molecular variance among individuals within and between phenotype groups was visualized in a factorial correspondence analysis performed over individuals in Genetix 4.05 (ref. 48).  $F_{ST}$  estimates and their significance were calculated over 100 permutations, as implemented in Arlequin<sup>46</sup>.

**Population genetics of opsin genes.** Determination of the *LWS* gene was as described previously<sup>13</sup>. We determined the sequences of exons 2–5 of *LWS* (872 bp), which encode the transmembrane region, from 263 individuals (526 haplotypes): Marumbi (12 individuals; 24 haplotypes), Luanso (27; 54), Kissenda (62; 124), Python (90; 180) and Makobe (72; 144). Additionally, we sequenced exons 2–5 of several hundred individuals of other species of Lake Victoria cichlids (Supplementary Fig. 4). Determination of the *SWS2A* gene is described in Supplementary Methods. We sequenced exons 1–5 (including introns) from males of Makobe (16 *P. pundamilia* and 17 *P. nyererei*) and Kissenda (20 blue and 20 red males).  $F_{ST}$  values for *LWS* and *SWS2A* sequences were calculated using DnaSP 4.0 (ref. 49). The *SWS2A* sequence (1,930 bp) was split into two putative alleles for the analysis.

**Molecular signature of selection on *LWS*.** Determination of the *LWS* flanking sequences and the tests for detection of selection were performed as described previously<sup>9</sup> with minor modifications. The *LWS* gene and its 5 kb upstream and 3.5 kb downstream flanking sequences (total 10.5 kb) were amplified by long PCR<sup>9</sup> from 10 red and 9 blue males. To reflect the approximate frequencies of *LWS* alleles in the two phenotype populations, we included one heterozygous (H/P) individual of each nuptial colour. The McDonald test<sup>38</sup> was calculated with the recombination parameter set to 2, 4, 10, 32 and 1,000 replicates.

**Female mating preferences.** We conducted laboratory two-way mate choice assays as described elsewhere<sup>40</sup>. Each female was tested in at least 5 trials with 5 different male pairs. A  $G$ -test was used to compare the frequency distributions of mating preferences between islands.

45. Venables, W. N. & Ripley, B. D. Modern applied statistics with S. (Springer, 2002).

46. Excoffier, L., Laval, G. & Schneider, S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1, 47–50 (2005).

47. Rice, W. R. Analyzing tables of statistical tests. *Evolution* 43, 223–225 (1989).

48. Belkhir, K., Borsa, P. & Chikhi, L., Raufaste, N. & Bonhomme, F. Genetix Version 4.05 for Windows Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France) (1996–2004); <http://www.genetix.univ-montp2.fr/genetix/genetix.htm>.

49. Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X., Rozas, R. & Dna, S. P. DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496–2497 (2003).



# In vivo reprogramming of adult pancreatic exocrine cells to $\beta$ -cells

Qiao Zhou<sup>1</sup>, Juliana Brown<sup>2</sup>, Andrew Kanarek<sup>1</sup>, Jayaraj Rajagopal<sup>1</sup> & Douglas A. Melton<sup>1</sup>

**One goal of regenerative medicine is to instructively convert adult cells into other cell types for tissue repair and regeneration. Although isolated examples of adult cell reprogramming are known, there is no general understanding of how to turn one cell type into another in a controlled manner. Here, using a strategy of re-expressing key developmental regulators *in vivo*, we identify a specific combination of three transcription factors (*Ngn3* (also known as *Neurog3*) *Pdx1* and *Mafa*) that reprograms differentiated pancreatic exocrine cells in adult mice into cells that closely resemble  $\beta$ -cells. The induced  $\beta$ -cells are indistinguishable from endogenous islet  $\beta$ -cells in size, shape and ultrastructure. They express genes essential for  $\beta$ -cell function and can ameliorate hyperglycaemia by remodelling local vasculature and secreting insulin. This study provides an example of cellular reprogramming using defined factors in an adult organ and suggests a general paradigm for directing cell reprogramming without reversion to a pluripotent stem cell state.**

Cells of adult organisms arise from sequential differentiation steps that are generally thought to be irreversible<sup>1</sup>. Biologists often describe this process of development as proceeding from an undifferentiated (embryonic) cell to a terminally differentiated cell that forms part of an adult tissue or organ. There are rare examples, however, in which cells of one type can be converted to another type in a process called cellular reprogramming or lineage reprogramming<sup>2,3</sup>. Various forms of cellular reprogramming are referred to in the literature as transdifferentiation, dedifferentiation or transdetermination<sup>4</sup>. For example, cellular reprogramming occurs in amphibian limb regeneration and fly imaginal disc identity switches<sup>5,6</sup>, and it may be central to certain types of pathological metaplasia<sup>4</sup>. There is long-standing interest and fascination in reprogramming studies, in part because of the promise of harnessing this phenomenon for regenerative medicine whereby abundant adult cells that can be easily collected would be converted to other medically important cell types to repair diseased or damaged tissues.

Somatic cell nuclear transfer (SCNT), developed in the 1960s, demonstrated that nuclei from differentiated adult cells could be reprogrammed to a totipotent state after injection into enucleated eggs<sup>2,7</sup>. More recently, it was shown that a small number of transcription factors can reprogram cultured adult skin cells to induced pluripotent stem (iPS) cells<sup>8–13</sup>. These studies point to the possibility of regenerating mammalian tissues by first reverting skin or other adult cells to pluripotent stem cells and then redifferentiating these into various cell types. Alternatively, it should be possible to convert one cell type into another directly, without the need to first revert the cell to an undifferentiated pluripotent state. Indeed, there are examples in the literature that suggest that this approach is feasible. For example, studies with embryonic cells have shown that dermal fibroblasts and retinal epithelial cells can be converted into muscle-like cells<sup>14</sup>, and pancreatic tissue to liver<sup>15</sup>. In adult animals, mature B lymphocytes have been reprogrammed into macrophages<sup>16</sup> or pro-B cells<sup>17</sup>. Today, well documented examples of cellular reprogramming, especially in adult animals, remain rare and have generally been restricted to cases in which a single inducing factor is involved. The recent work on iPS formation suggests that a specific combination of multiple factors, instead of a single one, might be the most effective way to reprogram adult cells<sup>8–13</sup>.

We developed a strategy to identify adult cell reprogramming factors by re-expressing multiple embryonic genes in living adult animals. Our focus on embryonic genes is based in part on regeneration studies in newts, frogs and fish, wherein it has been shown that dedifferentiation of adult cells to progenitors, a form of cellular reprogramming, is accompanied by reactivation of embryonic regulators<sup>5,18,19</sup>. These studies suggest that re-expression of appropriate embryonic genes may reprogram differentiated cells.

To search for factors that could reprogram adult cells into  $\beta$ -cells, we focused on transcription factors, a class of genes enriched for factors that regulate cell fates during embryogenesis. An *in situ* hybridization screen of more than 1,100 transcription factors identified groups of transcription factors with cell-type-specific expressions in the embryonic pancreas<sup>20</sup>. There are at least 20 transcription factors expressed in mature  $\beta$ -cells and their immediate precursors, the endocrine progenitors (Supplementary Table 1). Of these, 9 genes exhibited  $\beta$ -cell developmental phenotypes when mutated<sup>21,22</sup>, and these were selected for initial reprogramming experiments.

We chose mature exocrine cells of the adult pancreas as target cells for reprogramming. Exocrine cells derive from pancreatic endoderm, as do  $\beta$ -cells<sup>23</sup>, and exocrine cells can turn on endocrine programs when dissociated and cultured *in vitro*<sup>24,25</sup>. We carried out our experiments *in vivo* so that any induced  $\beta$ -cells would reside in their native environment, which might promote their survival and/or maturation. In addition, this approach allows for a direct comparison of endogenous and induced  $\beta$ -cells. The transcription factors were delivered into the pancreas in adenoviral vectors. It has been shown that adenovirus preferentially infects pancreatic exocrine cells, but not islet cells<sup>26</sup>, and, because most endogenous  $\beta$ -cells reside within islets (Fig. 1b), any newly formed (induced)  $\beta$ -cells could be easily detected as extra-islet insulin<sup>+</sup> cells.

## Induction of insulin<sup>+</sup> cells in adult mice

Adenovirus that co-expresses each transcription factor together with nuclear GFP (nGFP) was purified. All nine viruses were pooled and injected as a mixture (referred to as M9, for mixture of nine) into the pancreata of 2-month-old adult mice (Fig. 1a). The immune-deficient

<sup>1</sup>Department of Stem Cell and Regenerative Biology, Howard Hughes Medical Institute, Harvard Stem Cell Institute, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA. <sup>2</sup>Department of Pathology, Children's Hospital, Boston, Harvard Medical School, Harvard Stem Cell Institute, 300 Longwood Avenue, Boston, Massachusetts 02115-5724, USA.

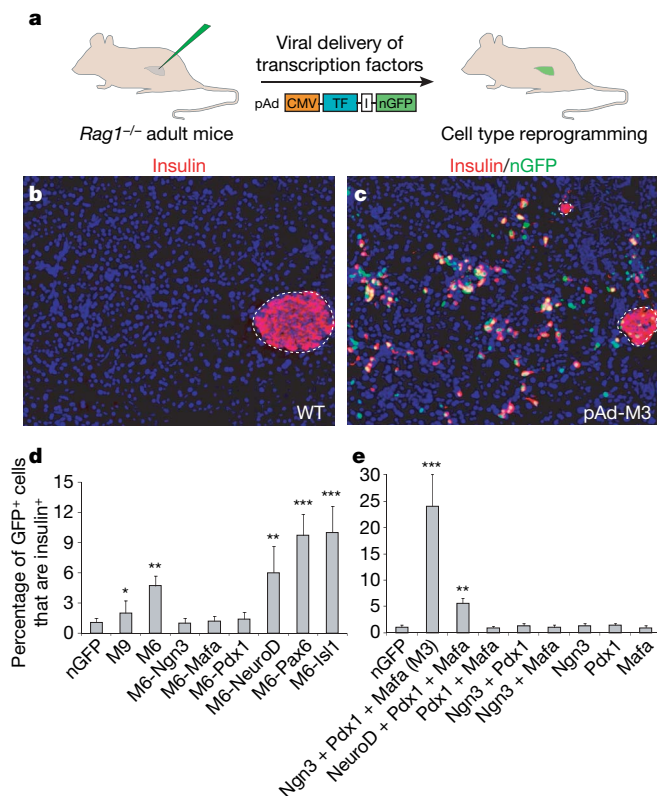
*Rag1*<sup>-/-</sup> strain was used to avoid complications associated with viral-elicited immune response<sup>27</sup>. One month after viral delivery, immunohistochemistry revealed a modest increase of extra-islet insulin<sup>+</sup> cells among viral infected cells (nGFP<sup>+</sup>) in two out of three animals (Fig. 1d). To determine which of the nine factors are required, individual factors were removed from the pool one at a time. Pools lacking Nkx2.2, Nkx6.1 or Pax4 continued to produce increased extra-islet insulin<sup>+</sup> cells (data not shown), suggesting that these genes are dispensable. Results for the other six genes were inconclusive. We conducted another round of factor withdrawal with mixtures of the remaining six genes (M6); three of them, *Ngn3*, *Pdx1* and *Mafa*, proved to be absolutely required (Fig. 1d). The combination of these three factors (referred to as M3) converted >20% of infected cells to insulin<sup>+</sup> cells (red cells with green nuclei, Fig. 1c, e). Notably, single factors or combinations of any two factors did not elicit this effect (Fig. 1e). Antibody labelling confirmed that these three inducing factors are co-expressed in the induced insulin<sup>+</sup> cells (Supplementary Fig. 1). NeuroD (also known as Neurod1) can functionally replace *Ngn3* in M3, but the resulting cocktail has reduced induction efficiency (Fig. 1e).

We noticed that the percentage of insulin<sup>+</sup> cells among infected cells increases with progressive removal of factors from the pool such that M3 induces more insulin<sup>+</sup> cells than M6, whereas M6 is better than M9 (Fig. 1d, e). This is probably due to the fact that a constant volume of virus was injected into each animal, regardless of the viral

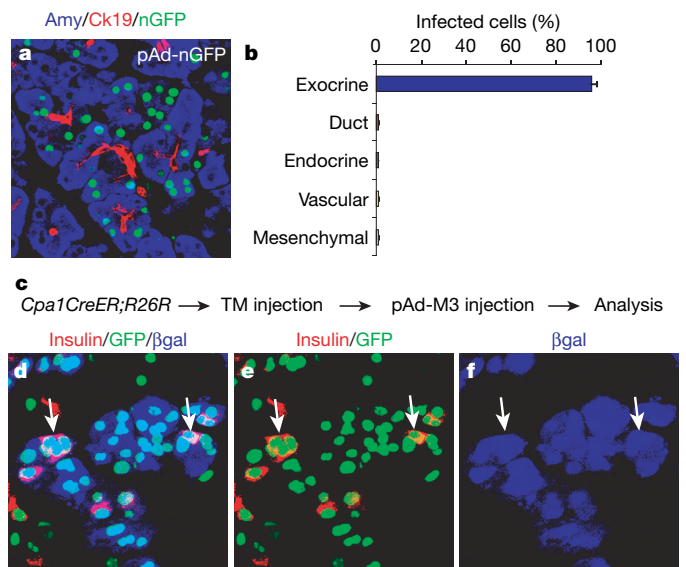
combinations. The effective concentration of *Ngn3*, *Pdx1* and *Mafa* viruses in a cocktail, therefore, increases when fewer factors are included. New insulin<sup>+</sup> cells were detected 3 days after injection, but the expression level was low. The intensity of insulin staining increased gradually so that, by day 10, the level was comparable to that of endogenous  $\beta$ -cells (Supplementary Fig. 2). These new insulin<sup>+</sup> cells were still present after 3 months, the longest time point that we analysed, and remained as scattered individual cells or small clusters and did not form islets (Fig. 1c). The reprogramming effect of the three factors appeared to be rather specific for pancreatic exocrine cells: infection of skeletal muscle *in vivo* or fibroblasts *in vitro* with M3 did not induce insulin expression, despite extensive co-expression of the three factors in the target cells (Supplementary Fig. 1).

### New insulin<sup>+</sup> cells come from exocrine cells

Lineage analysis was performed to determine the origin of the new insulin<sup>+</sup> cells. The five major cell types in the adult pancreas can be detected with lineage-specific molecular markers: exocrine (amylase), duct (Ck19), endocrine (insulin, glucagon, somatostatin and pancreatic polypeptide), vascular (PECAM) and mesenchymal (nestin and vimentin) cells. On injection with a control nGFP virus, most infected cells (>95%) were found to be mature amylase<sup>+</sup> exocrine cells (Fig. 2a, b), consistent with previous reports<sup>26</sup>. Non-exocrine cells together accounted for approximately 5% of the infected population. Because more than 20% of M3-infected cells become insulin<sup>+</sup> 10 days after viral delivery, it suggests that non-exocrine cells can contribute, at most, to a minor fraction of these new insulin<sup>+</sup> cells. As there is little cell death and no enhanced proliferation during this reprogramming (Supplementary Fig. 3), most insulin<sup>+</sup> cells would thus appear to originate from mature exocrine cells. To confirm the exocrine origin of the new insulin<sup>+</sup> cells, we genetically labelled mature exocrine cells with a mouse line (*Cpa1CreER*<sup>T2</sup>) that expresses an inducible form of Cre recombinase (*CreER*<sup>T2</sup>) specifically in adult exocrine cells<sup>20</sup> (Fig. 2c). When crossed with the *R26R* reporter line, tamoxifen induction in double heterozygous *Cpa1CreER*<sup>T2</sup>; *R26R* adults indelibly labelled 5–10% of mature



**Figure 1 | A combination of three transcription factors induces insulin<sup>+</sup> cells in adult mouse pancreas *in vivo*.** **a**, Schematic diagram of the experimental strategy. Adenoviruses encoding bicistronic transcription factor (TF) and nGFP linked by an IRES element (I) were injected into the pancreas of an adult mouse (*Rag1*<sup>-/-</sup>). CMV, cytomegaloviral promoter. **b**, Wild type (WT) pancreas is predominantly exocrine tissue with insulin<sup>+</sup>  $\beta$ -cells in the islet (outlined). Nuclei were stained blue with DAPI. **c**, One month after infection with a combination of *Ngn3*, *Pdx1* and *Mafa* viruses (pAd-M3), numerous insulin<sup>+</sup> cells appear outside of islets. **d, e**, Quantification of induction one month after infection. M9, M6: mixture of 9 and 6 different viruses, respectively. Data are presented as mean + s.d.; *n* = 3 animals. ~1,000 nGFP<sup>+</sup> cells were counted per animal. Asterisk, *P* < 0.05; two asterisks, *P* < 0.01; three asterisks, *P* < 0.001.



**Figure 2 | Induced new  $\beta$ -cells originate from differentiated exocrine cells.** **a**, Ten days after nGFP viral infection, most infected cells are amylase<sup>+</sup> (*Amy*<sup>+</sup>) mature exocrine cells, not duct cells (*Ck19*<sup>+</sup>). **b**, Quantification of nGFP-infected cell types. Data are presented as mean + s.d., *n* = 3 animals. ~1,000 nGFP<sup>+</sup> cells were counted. **c**, Double heterozygous *Cpa1CreER*<sup>T2</sup>; *R26R* adult mice are injected with tamoxifen (TM), which labels the mature exocrine cells with  $\beta$ -galactosidase ( $\beta$ gal). Reprogramming is subsequently induced by infection with pAd-M3. **d–f**, Ten days after infection, many  $\beta$ gal<sup>+</sup> insulin<sup>+</sup> cells (arrows) are present. **e** and **f** are insulin (red)/GFP (green) and  $\beta$ gal (blue) channels of **d**, respectively.

exocrine cells with  $\beta$ -galactosidase (Fig. 2d, f); no label was found in other cell types. After pAd-M3 injection, many  $\beta$ -galactosidase<sup>+</sup> cells become insulin<sup>+</sup> (Fig. 2d–f, pink cells), providing direct evidence that mature exocrine cells give rise to new insulin<sup>+</sup> cells.

### Induced $\beta$ -cells closely resemble islet $\beta$ -cells

We next examined the new insulin<sup>+</sup> cells to determine the extent to which they have been reprogrammed. Morphologically, exocrine cells are large with a cobble stone appearance (Fig. 3a, b) whereas islet  $\beta$ -cells are much smaller and spindle shaped (Fig. 3a). When dissociated into single cells, the diameter of amylase<sup>+</sup> exocrine cells range from 25  $\mu$ m to 17  $\mu$ m whereas insulin<sup>+</sup>  $\beta$ -cells range from 9  $\mu$ m to 15  $\mu$ m. The induced cells are indistinguishable from islet  $\beta$ -cells in size and shape (Fig. 3b, c).

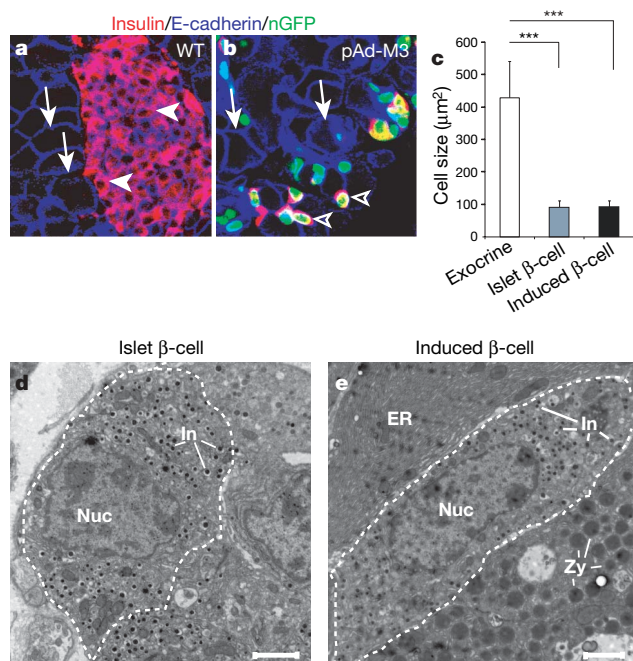
At the ultrastructural level, the reprogrammed cells have all the hallmarks of islet  $\beta$ -cells (Fig. 3d, e). They possess the small dense secretory granules characteristic of insulin granules, and lack the large zymogen granules and dense assemblies of endoplasmic reticulum that are characteristic of exocrine cells (Fig. 3d, e). Immunoelectron microscopy further showed that the induced  $\beta$ -cells express both GFP in the nucleus and abundant insulin in the granules (Supplementary Fig. 4). Interestingly, the induced  $\beta$ -cells often appeared on the electron micrograph as intercalated within exocrine acinar rosettes (Fig. 3e). In wild-type pancreatic samples, rare single or small clusters of  $\beta$ -cells reside outside islets, but they often associate with duct but not exocrine cells. The unique position of induced cells probably reflects their exocrine origin.

Molecular marker analysis reveals that most of the insulin<sup>+</sup> cells co-express genes essential for  $\beta$ -cell endocrine function including glucose transporter 2 (Glut2, also known as Slc2a2, expressed in

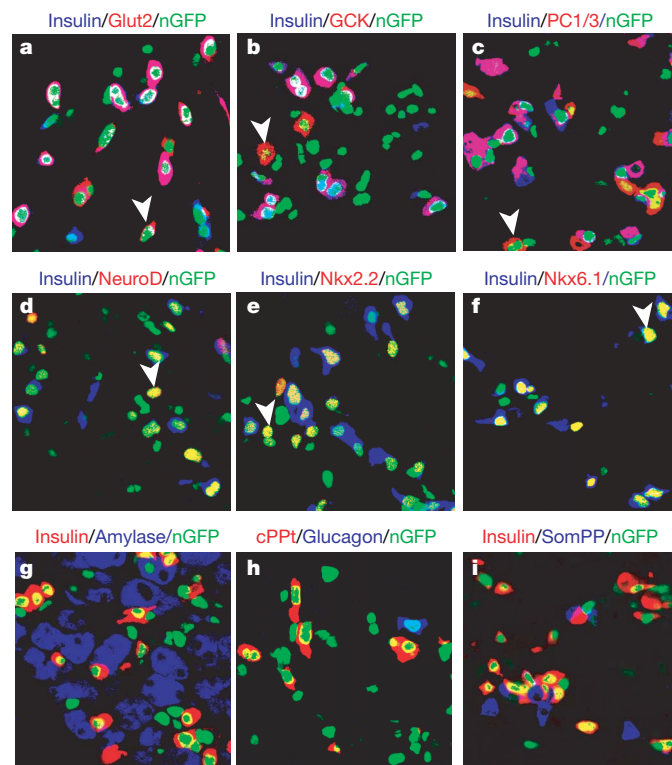
92.8% of the new insulin<sup>+</sup> cells), glucokinase (GCK, 96.7%), pro-hormone convertase (PC1/3, also known as Pcsk1, 86.7%; Fig. 4a–c and Supplementary Fig. 5), and the key  $\beta$ -cell transcription factors NeuroD (88.9%), Nkx2.2 (85.3%) and Nkx6.1 (85.9%; Fig. 4d–f and Supplementary Fig. 5). The induced insulin<sup>+</sup> cells express C-peptide (part of proinsulin; Fig. 4h). Expression profile analysis of the reprogrammed cells further indicates a strong overlap of endocrine-enriched genes between reprogrammed cells and islet cells, suggesting a high degree of similarity between their endocrine programs (Supplementary Fig. 6).

The new  $\beta$ -cells do not express exocrine genes such as amylase or Ptf1a, the duct marker Ck19 (also known as Krt19), mesenchymal markers nestin and vimentin, nor the neuronal marker Tuji ( $\beta$ -tubulin III, also known as Tubb3) (Fig. 4g, Supplementary Fig. 5 and data not shown). Nor do the new  $\beta$ -cells express any other pancreatic hormones such as glucagon, somatostatin or pancreatic polypeptide (Fig. 4h, i and Supplementary Fig. 5). Thus, the new  $\beta$ -cells do not exhibit a hybrid or mixed phenotype, indicating silencing of non- $\beta$ -cell programs.

The primary function of  $\beta$ -cells is to synthesize and release insulin. To facilitate the release of insulin into the circulation,  $\beta$ -cells, unique among pancreatic cell types, synthesize vascular endothelial growth factor (VEGF), which promotes local angiogenic remodelling<sup>28</sup>. Notably, induced  $\beta$ -cells similarly synthesize VEGF and induce angiogenesis so that blood vessels form next to these new cells (Fig. 5a, b). Quantification indicates that, in nGFP controls, 32% of infected cells lie adjacent to blood vessels whereas 61% and 83% of induced  $\beta$ -cells are directly juxtaposed to blood vessels 10 days and 30 days after induction, respectively.



**Figure 3 | Endogenous and induced  $\beta$ -cells are indistinguishable in morphology and ultrastructure.** **a, b**, Islet  $\beta$ -cells (**a**, arrowheads) and induced  $\beta$ -cells (**b**, arrowheads) are similar in size and shape but distinctly different from exocrine cells (**a, b**, arrows). E-cadherin staining was used to visualize cell boundaries. **c**, Size comparison of exocrine cells, islet  $\beta$ -cells and induced  $\beta$ -cells. Data are presented as mean  $\pm$  s.d.,  $n = 3$  animals.  $>100$  cells per animal were used. Three asterisks,  $P < 0.001$ . **d**, Electron micrograph of a  $\beta$ -cell (outlined) in an islet. **e**, Example of an induced  $\beta$ -cell situated between two exocrine cells. Endogenous and induced  $\beta$ -cells contain small insulin granules (In) and lack zymogen granules (Zy) of exocrine cells and extensive endoplasmic reticulum (ER). Nuc, nucleus. Scale bars, 2  $\mu$ m.



**Figure 4 | Molecular marker characterization of induced  $\beta$ -cells.** **a–f**, One month after infection with pAd-M3, most insulin<sup>+</sup> induced  $\beta$ -cells co-express endocrine genes: glucose transporter 2 (Glut2, **a**), glucokinase (GCK, **b**), prohormone convertase 1/3 (PC1/3, **c**) and  $\beta$ -cell transcription factors NeuroD (**d**), Nkx2.2 (**e**) and Nkx6.1 (**f**). Arrowheads indicate examples of cells that express marker genes but not insulin. **g–i**, Induced  $\beta$ -cells do not express amylase (**g**), glucagon (**h**) or somatostatin/pancreatic polypeptide (**i**, SomPP). cPpt, c-peptide.



To test whether induced  $\beta$ -cells release insulin, mice were rendered diabetic by streptozotocin (STZ) injection, which specifically ablates islet  $\beta$ -cells. When subsequently injected with pAd-M3, fasting blood glucose levels of hyperglycaemic animals showed a significant and long-lasting improvement compared to animals injected with control (nGFP) virus (Fig. 5d). In addition, the pAd-M3 animals showed increased glucose tolerance (Supplementary Fig. 7), had increased insulin levels in the serum (non-fasting,  $P < 0.01$ , Fig. 5e) and possessed large numbers of induced  $\beta$ -cells (Fig. 5f). Polymerase chain reaction with reverse transcription (RT-PCR) analysis and direct observation revealed that virus injected into the pancreas does not spread to other internal organs such as liver and intestine that, theoretically, could modulate insulin secretion and/or response (Supplementary Fig. 8). In addition, we found no evidence that STZ-treated animals show spontaneous conversion of exocrine cells to  $\beta$ -cells (Supplementary Fig. 8). As the data in Fig. 5 show, the total number of induced  $\beta$ -cells is rather small compared to the number of  $\beta$ -cells in normal animals and this may account for the limitation to the effectiveness in restoring glucose homeostasis. Alternatively, because the new  $\beta$ -cells are not reorganized into islet structures, this may limit their effectiveness. Together, these data show that induced  $\beta$ -cells can produce and secrete insulin *in vivo*.

### Inducing factors are required only transiently

Our results thus far support the contention that a combination of three transcription factors fully reprograms exocrine cells to  $\beta$ -cells *in vivo*. To determine whether continued presence of these factors is required to maintain the phenotype of reprogrammed cells, we used RT-PCR and primers specific to viral transgenes to detect their presence. Transgene expression from all three viruses was substantially diminished after 1 month and was undetectable after 2 months

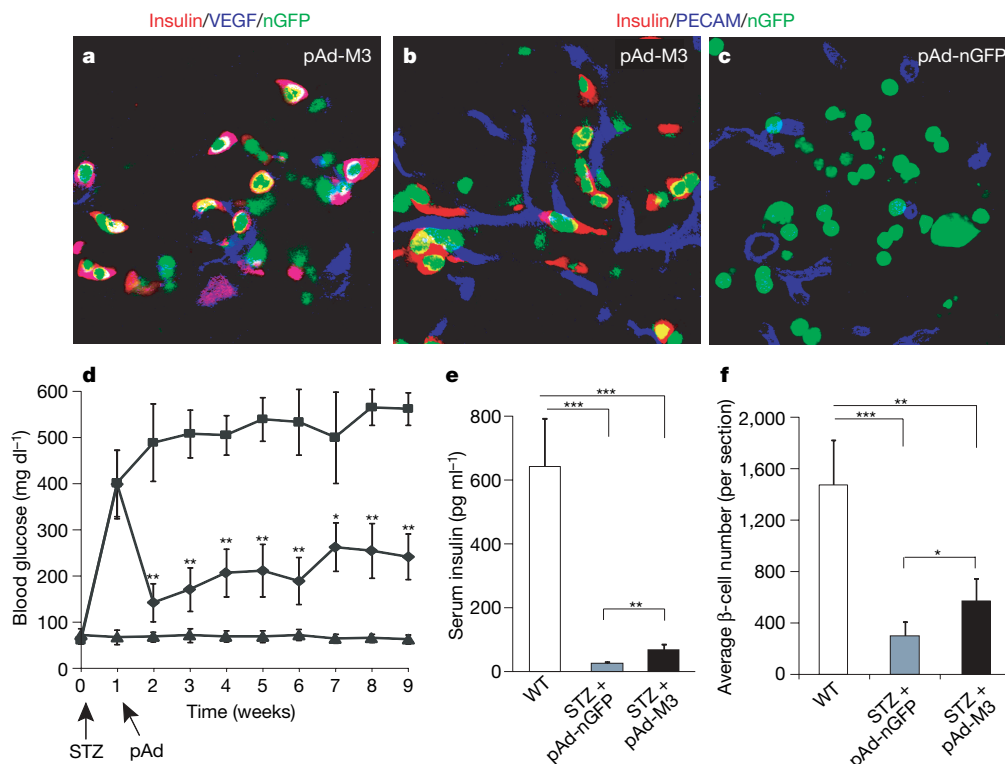
(Supplementary Fig. 9). Ngn3 protein was undetectable by antibody staining 1 month after infection (Supplementary Fig. 9). Pdx1 and Maf protein expression in the induced  $\beta$ -cells, however, remains consistently strong even after 2 months, indicating the activation of endogenous genes (Supplementary Fig. 9). These results are consistent with the fact that endogenous islet  $\beta$ -cells do not express Ngn3, but do express Pdx1 and Maf<sup>21,22</sup>. Thus, a transient expression of the inducing factors is sufficient to convert exocrine cells to a stable new  $\beta$ -cell state.

### $\beta$ -cell reprogramming does not involve dedifferentiation

In principle, the conversion of exocrine cells to  $\beta$ -cells could be direct or involve dedifferentiation to common progenitors that then redifferentiate into  $\beta$ -cells. Indeed, exocrine and  $\beta$ -cells share a common progenitor during embryogenesis that is characterized by rapid division and expression of genes including *Sox9* and *Hnf6* (also known as *Oncut1*; ref. 20). Continuous 5-bromodeoxyuridine (BrdU) labelling over the first 10 days of reprogramming, however, shows that few induced  $\beta$ -cells (3.2%) have divided (Supplementary Fig. 3). In comparison, 12.9% of endogenous islet  $\beta$ -cells in the same animals incorporated BrdU (Supplementary Fig. 3). In addition, we detected no induction of *Sox9* or *Hnf6* (data not shown). These results suggest that *in vivo* reprogramming of exocrine to  $\beta$ -cells is a direct conversion of cell types and does not involve dedifferentiation. We can not formally exclude the possibility that a very transient or partial dedifferentiation may occur, but our results indicate that extensive replication and reversion to a dedifferentiated cell for an appreciable time does not occur.

### Discussion

Our results provide evidence that fully differentiated exocrine cells can be directly reprogrammed into cells that closely resemble  $\beta$ -cells



**Figure 5 | Induced new  $\beta$ -cells remodel vasculature and ameliorate hyperglycaemia.** **a–c**, New  $\beta$ -cells synthesize VEGF (**a**) and induce local angiogenic remodelling (**b**). Note the proximity of blood vessels (PECAM<sup>+</sup>) to the induced  $\beta$ -cells (**b**) versus control infected cells (**c**). **d**, Improvement of fasting blood glucose level in diabetic mice after injection with pAd-M3 (diamond) compared to controls with nGFP virus (square). Triangle,

non-diabetic controls. STZ, streptozotocin. Arrows indicate timing of injection.  $n = 6$ –8 animals. **e**, Non-fasting serum insulin levels 6 weeks after injection.  $n = 6$ –8 animals. **f**, Average insulin<sup>+</sup>  $\beta$ -cell number per section 8 weeks after injection.  $n = 3$  animals. Both islet and induced  $\beta$ -cells were counted for the pAd-M3 samples. One asterisk,  $P < 0.05$ ; two asterisks,  $P < 0.01$ ; three asterisks,  $P < 0.001$ . Data are presented as mean  $\pm$  s.d.

in adult animals by a combination of just three transcription factors. The three reprogramming factors, *Ngn3*, *Pdx1* and *Mafa*, are known to be important in the embryonic development of pancreas and  $\beta$ -cells<sup>21,22</sup>. In contrast, many additional factors are also required for  $\beta$ -cell development<sup>21,22</sup>. Further studies will be necessary to understand why this particular combination is sufficient for adult  $\beta$ -cell reprogramming.

The induced  $\beta$ -cells do not organize into islet structures and remain as single cells or small clusters. Signalling between  $\beta$ -cells inhibits basal insulin secretion and enhances glucose-stimulated insulin secretion<sup>29</sup>. The lack of organization of induced  $\beta$ -cells undoubtedly impairs their function. Strategies that promote aggregation of the induced  $\beta$ -cells in adult should help to restore full glucose responsiveness.

There have been previous attempts to convert adult liver cells to  $\beta$ -cells *in vivo* by expressing pancreatic transcription factors<sup>27,30–32</sup>. These factors were able to induce expression of some pancreatic genes, but not phenotypic or morphological conversion into functional  $\beta$ -cells<sup>27,30–32</sup>. Mature exocrine cells can turn on endocrine programs when dissociated and cultured *in vitro*<sup>24,25,33,34</sup>. Interestingly, dissociation itself is apparently sufficient to initiate endocrine programs whereas the addition of growth factors is necessary for cell survival<sup>24,25,33,34</sup>. However, the molecular mechanisms of this process remain largely unknown. Other studies have shown that pancreatic duct cells and liver cells could be induced to express certain  $\beta$ -cell gene products in culture<sup>35–37</sup>. Most of these studies, however, did not address whether these cells possess a hybrid phenotype. In addition, RT-PCR on populations of cells, instead of single-cell resolution immunohistochemistry, was routinely used to evaluate the expression of  $\beta$ -cells markers. It is unclear how many cells actually expressed these markers or at what level. Finally,  $\beta$ -cells exhibit highly unstable phenotypes when cultured and appear to transform into fibroblast-like cells<sup>38,39</sup>. *In vitro* generation of  $\beta$ -cells will probably require suitable culture conditions that have yet to be discovered.

It is surprising that the reprogramming of exocrine cells to  $\beta$ -cells does not involve multiple rounds of cell proliferation. It is generally thought that epigenetic changes that underlie reprogramming events are most easily made during cell division<sup>2</sup>. It may be the case that many reprogramming events do indeed involve obligatory proliferation steps<sup>4</sup>. In contrast, reprogramming of B lymphocytes to macrophages seems to be cell-cycle-independent<sup>16</sup>. Early SCNT experiments also provided evidence for reprogramming without DNA replication<sup>40</sup>.

Reprogramming of exocrine cells to  $\beta$ -cells occurs at a relatively fast speed, with the first insulin<sup>+</sup> cells appearing at day 3, and with efficiency of up to 20%. This is in contrast with recent reports of reprogramming fibroblasts to embryonic stem cells<sup>8–13</sup>, where it takes a considerably longer time (7–30 days) and the efficiency is much reduced (typically less than 0.1%). This may be due to the fact that pancreatic exocrine and  $\beta$ -cells are closely related cell types and share much of their epigenomes whereas the epigenomes of fibroblasts and embryonic stem cells are largely dissimilar. Conversion between exocrine and  $\beta$ -cells may therefore require fewer epigenetic changes.

Recent advances in mammalian cellular reprogramming with defined genes collectively point to the possibility that a limited number of factors could reprogram any given adult cell to a different type of cell such as a stem cell, a committed progenitor or another mature cell type. All these studies relied on knowledge of the normal development of these cell types, which enabled the manipulation of key developmental regulators in adult cells. This approach may prove to be a general strategy for directing adult cell reprogramming. The recent reprogramming of human skin cells to iPS cells raises the possibility of generating patient-specific human embryonic stem lines for therapies<sup>9,10,13</sup>. This would be the first step in a process that will then require directed differentiation of the iPS cells to produce therapeutically important cell types such as neurons, cardiomyocytes or pancreatic  $\beta$ -cells. In principle, patient-specific cell therapies

could be achieved more directly by reprogramming abundant and easily accessible patient-specific human cells such as fibroblasts, blood cells or adipocytes.

## METHODS SUMMARY

**Adenovirus construction and purification.** Genes of interest were first cloned into a shuttle vector containing an internal ribosome entry site linked to nuclear GFP (*IRES-nGFP*), and then into the pAd/CMV/V5-DEST adenoviral vector (Invitrogen). High titre virus ( $>1 \times 10^{10}$  plaque-forming units (p.f.u.) per ml) was obtained by purification with the AdEasy Kit (Stratagene).

**Animals, surgery and physiological studies.** *Rag1*<sup>−/−</sup> and *Rag1*<sup>−/−</sup>; *NOD* animals were obtained from Jackson Laboratories. Adult animals ( $>2$  months old) were injected with 100  $\mu$ l ( $>1 \times 10^9$  p.f.u.) of purified adenovirus directly into the splenic lobe of the dorsal pancreas. Blood glucose was measured with Ascensia Elite blood glucose meter. Insulin levels were determined with an Ultrasensitive insulin ELISA kit (Alpco).

**Immunohistochemistry, BrdU labelling and TUNEL analysis.** This was performed as previously described<sup>41</sup>. BrdU (1 mg ml<sup>−1</sup>) was provided in drinking water for BrdU labelling after surgery. Apoptotic cells were recognized by TUNEL (terminal dUTP nick-end labelling) with a TMR red cell death kit (Roche).

**Electron microscopy.** Dissected pancreas was fixed in 4% paraformaldehyde and 0.1% glutaraldehyde for 2 h at room temperature (24 °C). For conventional transmission electron microscopy, samples were further fixed by osmium tetroxide, embedded in Epon resin and sectioned at 60–80 nm. For immunogold labelling, ultrathin sections were cut at  $-120$  °C and stained with gold-conjugated antibodies. Images were obtained with a Tecnai G<sup>2</sup> Spirit BioTWIN transmission electron microscope.

**FACS analysis and gene profiling.** Pancreas was digested with liberase and elastase (Roche) to single cells. GFP<sup>+</sup> cells were isolated by FACS with FACSaria (BD Bioscience). Biotin-labelled complementary RNA probes were synthesized with the Illumina TotalPrep RNA Amplification kit (Ambion). Gene profiling was performed with Sentrix BeadChip Array MouseRef-8 v1.1 (Illumina). Data were analysed with the BeadStudio software.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 26 June; accepted 6 August 2008.

Published online 27 August 2008.

- Weissman, I. L. Stem cells: units of development, units of regeneration, and units in evolution. *Cell* **100**, 157–168 (2000).
- Hochedlinger, K. & Jaenisch, R. Nuclear reprogramming and pluripotency. *Nature* **441**, 1061–1067 (2006).
- Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
- Slack, J. M. Metaplasia and transdifferentiation: from pure biology to the clinic. *Nature Rev. Mol. Cell Biol.* **8**, 369–378 (2007).
- Brockes, J. P. & Kumar, A. Plasticity and reprogramming of differentiated cells in amphibian regeneration. *Nature Rev. Mol. Cell Biol.* **3**, 566–574 (2002).
- Hadorn, E. Transdetermination in cells. *Sci. Am.* **219**, 110–114 (1968).
- Gurdon, J. B. From nuclear transfer to nuclear reprogramming: the reversal of cell differentiation. *Annu. Rev. Cell Dev. Biol.* **22**, 1–22 (2006).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
- Meissner, A., Wernig, M. & Jaenisch, R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nature Biotechnol.* **25**, 1177–1181 (2007).
- Wernig, M. *et al.* *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
- Park, I. H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141–146 (2008).
- Choi, J. *et al.* MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. *Proc. Natl Acad. Sci. USA* **87**, 7988–7992 (1990).
- Shen, C. N., Slack, J. M. & Tosh, D. Molecular basis of transdifferentiation of pancreas to liver. *Nature Cell Biol.* **2**, 879–887 (2000).
- Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663–676 (2004).
- Cobaleda, C., Jochum, W. & Busslinger, M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* **449**, 473–477 (2007).

18. Whitehead, G. G., Makino, S., Lien, C. L. & Keating, M. T. *fgf20* is essential for initiating zebrafish fin regeneration. *Science* **310**, 1957–1960 (2005).
19. Tanaka, E. M. Cell differentiation and cell fate during urodele tail and limb regeneration. *Curr. Opin. Genet. Dev.* **13**, 497–501 (2003).
20. Zhou, Q. *et al.* A multipotent progenitor domain guides pancreatic organogenesis. *Dev. Cell* **13**, 103–114 (2007).
21. Murtaugh, L. C. & Melton, D. A. Genes, signals, and lineages in pancreas development. *Annu. Rev. Cell Dev. Biol.* **19**, 71–89 (2003).
22. Jensen, J. Gene regulatory factors in pancreatic development. *Dev. Dyn.* **229**, 176–200 (2004).
23. Gu, G., Dubauskaite, J. & Melton, D. A. Direct evidence for the pancreatic lineage: NGN3<sup>+</sup> cells are islet progenitors and are distinct from duct progenitors. *Development* **129**, 2447–2457 (2002).
24. Baeyens, L. *et al.* *In vitro* generation of insulin-producing beta cells from adult exocrine pancreatic cells. *Diabetologia* **48**, 49–57 (2005).
25. Minami, K. *et al.* Lineage tracing and characterization of insulin-secreting cells generated from adult pancreatic acinar cells. *Proc. Natl Acad. Sci. USA* **102**, 15116–15121 (2005).
26. Wang, A. Y., Peng, P. D., Ehrhardt, A., Storm, T. A. & Kay, M. A. Comparison of adenoviral and adeno-associated viral vectors for pancreatic gene delivery *in vivo*. *Hum. Gene Ther.* **15**, 405–413 (2004).
27. Wang, A. Y., Ehrhardt, A., Xu, H. & Kay, M. A. Adenovirus transduction is required for the correction of diabetes using Pdx-1 or Neurogenin-3 in the liver. *Mol. Ther.* **15**, 255–263 (2007).
28. Lammert, E. *et al.* Role of VEGF-A in vascularization of pancreatic islets. *Curr. Biol.* **13**, 1070–1074 (2003).
29. Konstantinova, I. *et al.* EphA-Ephrin-A-mediated beta cell communication regulates insulin secretion from pancreatic islets. *Cell* **129**, 359–370 (2007).
30. Ferber, S. *et al.* Pancreatic and duodenal homeobox gene 1 induces expression of insulin genes in liver and ameliorates streptozotocin-induced hyperglycemia. *Nature Med.* **6**, 568–572 (2000).
31. Kaneto, H. *et al.* PDX-1/VP16 fusion protein, together with NeuroD or Ngn3, markedly induces insulin gene transcription and ameliorates glucose tolerance. *Diabetes* **54**, 1009–1022 (2005).
32. Miyatsuka, T. *et al.* Ectopically expressed PDX-1 in liver initiates endocrine and exocrine pancreas differentiation but causes dysmorphogenesis. *Biochem. Biophys. Res. Commun.* **310**, 1017–1025 (2003).
33. Minami, K. & Seino, S. Pancreatic acinar-to-beta cell transdifferentiation *in vitro*. *Front. Biosci.* **13**, 5824–5837 (2008).
34. Okuno, M. *et al.* Generation of insulin-secreting cells from pancreatic acinar cells of animal models of type 1 diabetes. *Am. J. Physiol. Endocrinol. Metab.* **292**, E158–E165 (2007).
35. Sapir, T. *et al.* Cell-replacement therapy for diabetes: generating functional insulin-producing tissue from adult human liver cells. *Proc. Natl Acad. Sci. USA* **102**, 7964–7969 (2005).
36. Heremans, Y. *et al.* Recapitulation of embryonic neuroendocrine differentiation in adult human pancreatic duct cells expressing neurogenin 3. *J. Cell Biol.* **159**, 303–312 (2002).
37. Gasa, R. *et al.* Proendocrine genes coordinate the pancreatic islet differentiation program *in vitro*. *Proc. Natl Acad. Sci. USA* **101**, 13245–13250 (2004).
38. Morton, R. A., Geras-Raaka, E., Wilson, L. M., Raaka, B. M. & Gershengorn, M. C. Endocrine precursor cells from mouse islets are not generated by epithelial-to-mesenchymal transition of mature beta cells. *Mol. Cell. Endocrinol.* **270**, 87–93 (2007).
39. Gershengorn, M. C. *et al.* Epithelial-to-mesenchymal transition generates proliferative human islet precursor cells. *Science* **306**, 2261–2264 (2004).
40. De Robertis, E. M. & Gurdon, J. B. Gene activation in somatic nuclei after injection into amphibian oocytes. *Proc. Natl Acad. Sci. USA* **74**, 2470–2474 (1977).
41. Dor, Y., Brown, J., Martinez, O. I. & Melton, D. A. Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. *Nature* **429**, 41–46 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to M. Ericsson for expert assistance on electron microscopy, R. Hellmiss-Peralta for advice on graphics, and B. Tilton and P. Rogers for FACS. We thank R. Martinez and G. Kenty for technical assistance; H. Edlund for the gift of Ptf1a antiserum; A. Kweudjeu for microarray analysis; members of the Melton laboratory for advice and feedback; and J. Sneddon, J. Annes and W. Anderson for critical reading of the manuscript. Q.Z. was supported by a Damon-Runyon Cancer Research Foundation Postdoctoral Fellowship and a Pathway to Independence (PI) Award from the National Institute of Health. D.A.M. is an HHMI investigator and this work was supported in part by the Harvard Stem Cell Institute and the NIH.

**Author Information** The microarray data were deposited in the Gene Expression Omnibus (GEO) under accession number GSE12025. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.A.M. ([dmelton@harvard.edu](mailto:dmelton@harvard.edu)).



## METHODS

**Viral injection and tissue collection.** For adult pancreas, ~100 µl purified virus was injected directly into 2–3 foci of the dorsal splenic lobe with a 3/10 cc Insulin Syringe (Becton Dickinson). For skeletal muscle, ~20 µl virus was injected into the upper thigh. At the time of tissue collection, the infected portion of the tissue was visualized by GFP fluorescence and dissected out. For adult pancreas, typically ~50% of the dorsal pancreas was taken.

**Immunohistochemistry.** Adult mouse pancreata were fixed by immersion in 4% paraformaldehyde for 2 h at 4 °C. Samples were subsequently incubated in 30% sucrose solution overnight (6–12 h) and embedded with optimal cutting temperature compound (Tissue-Tek).

The following primary antibodies were used: rat anti-E-cadherin (Zymed), rat anti-Pecam1 (Pharmingen), goat anti-Ngn3 (Santa Cruz), guinea-pig anti-insulin (Dako), guinea-pig anti-glucagon (Linco), guinea-pig anti-Pancreatic polypeptide (Linco), rabbit anti-somatostatin (Dako), rabbit anti-pancreatic polypeptide (Dako), goat anti-somatostatin (Santa Cruz), goat anti-Pdx1 (Santa Cruz), guinea-pig anti-Pdx1 (gift from C. Wright), goat anti-β-galactosidase (Biogenesis), goat anti-amylase (Santa Cruz), mouse anti-BrdU (Amersham), rabbit anti-mafa (Bethyl), chick anti-nestin (Ames), chick anti-ventralin (Chemicon), goat anti-Glut2 (Santa Cruz), goat anti-VEGF (R&D), rabbit anti-PC1/3 (Chemicon), goat anti-glucokinase (Santa Cruz), rabbit anti-Ck19 (Melton laboratory stock), rabbit anti-chromogranin A/B (RDI), rabbit anti-Ptfla (gift from H. Edlund), goat anti-NeuroD (Santa Cruz), rabbit anti-Nkx6.1 (BCBC), rabbit anti-Sox9 (Santa Cruz), goat anti-Nkx2.2 (Santa Cruz) and rabbit anti-c-peptide (Linco).

Rodamin-Red-X, FITC-, Cy5- and Alexa-dye-conjugated donkey secondary antibodies were obtained from the Jackson ImmunoResearch Laboratories and Molecular Probes Inc. Tyramide amplification system (PerkinElmer) was used for PC1/3 and glucokinase staining. Immunofluorescence pictures were taken with a Zeiss LSM 510 META confocal microscope.

***Cpa1CreER<sup>T2</sup>* labelling of mature exocrine cells.** *Cpa1CreER<sup>T2</sup>;R26R* double heterozygous animals were generated by mating homozygous *Cpa1CreER<sup>T2</sup>* males with *R26R* homozygous females (Jackson laboratory). Two-month-old *Cpa1CreER<sup>T2</sup>;R26R* adults were injected with tamoxifen at 6 mg per animal every third day four times to label mature exocrine cells.

**Physiological studies.** Diabetic animals were produced with intraperitoneal injection of streptozotocin (120 µg per g body weight) after overnight fasting with 2-month-old adult animals of the Rag1 strain (Jackson laboratory). Hyperglycaemic animals that displayed >250 mg dl<sup>-1</sup> fasting blood glucose levels for at least two consecutive days were used for experiments.

Fasting blood glucose was measured on tail-vein blood with an Ascensia Elite glucometer (Bayer) after 6–8 h fasting. The non-fasting insulin level was determined from tail-vein blood collected around 9 to 10 am with an Ultrasensitive Insulin ELISA kit (Alpco).

The average β-cell number per section was determined by sectioning through the entire pancreas at 15 µm and collecting every third section. Twenty randomly selected sections were immunostained for insulin and 4,6-diamidino-2-phenylindole (DAPI) to visualize individual β-cells. The total number of β-cells was counted and averaged from three animals.

The glucose tolerance test was performed by fasting animals overnight (12 h), followed by intraperitoneal injection of glucose (3 g per kg body weight).

**Electron microscopy.** Small pieces of pancreatic samples (1–2 mm) were fixed with 4% paraformaldehyde and 0.1% glutaraldehyde for 2 h at room temperature.

For conventional electron microscopy, samples were further refixed with a mixture of 1% osmium tetroxide (OsO<sub>4</sub>) plus 1.5% potassium ferrocyanide (K<sub>4</sub>Fe(CN)<sub>6</sub>) for 2 h, were washed in water and stained in 1% aqueous uranyl acetate for 1 h followed by dehydration in grades of alcohol (50%, 70%, 95%, 2 × 100%) and propyleneoxide (1 h), and then infiltrated in propyleneoxide:Epon 1:1 overnight and embedded in TAAB Epon (Marivac

Canada Inc.). Ultrathin sections (about 60–80 nm) were cut on a Reichert Ultracut-S microtome, picked up on to copper grids, stained with 0.2% lead citrate and examined in a Tecnai G<sup>2</sup> Spirit BioTWIN transmission electron microscope. Images were taken with an AMT CCD camera.

For immunoelectron microscopy, fixed samples were infiltrated with 2.3 M sucrose in PBS for 30 min then frozen in liquid nitrogen. Frozen samples were sectioned at –120 °C, the sections transferred to formvar–carbon-coated copper grids and floated on PBS until the immunogold labelling was carried out.

The gold labelling was carried out at room temperature on a piece of paraffin. All antibodies and protein-A gold were diluted in 1% BSA. The diluted antibody solution was centrifuged for 1 min at >10,000g before labelling to avoid possible aggregates. Grids were floated on drops of 1% BSA for 10 min to block unspecific labelling, transferred to 5-µl drops of primary antibody and incubated for 30 min. The grids were then washed in four drops of PBS for a total of 15 min, transferred to 5-µl drops of protein-A gold (G. Posthuma) for 20 min, and washed in four drops of PBS for 15 min and six drops of double-distilled water.

For double labelling, after the first protein-A gold incubation, grids were washed in four drops of PBS for a total of 15 min and then transferred to a drop of 0.2% glutaraldehyde in PBS for 5 min, and washed in four drops of PBS/0.15 M glycine (to quench free aldehyde groups). Following this, the second primary antibody was applied, followed by PBS wash and different size protein-A gold as described previously. The antibodies used were rabbit anti-GFP (Invitrogen) and guinea-pig anti-insulin (Dako).

Contrasting/embedding of the labelled grids was carried out on ice in 0.3% uranyl acetate (Electron Microscopy Sciences) in 2% methyl cellulose (Sigma) for 10 min. Grids were picked up with metal loops (diameter slightly larger than the grid) and the excess liquid was removed by streaking on a filter paper (Whatman, number 1), leaving a thin coat of methyl cellulose (bluish interference colour when dry).

The grids were examined in a Tecnai G<sup>2</sup> Spirit BioTWIN transmission electron microscope and images were recorded with a 2k AMT CCD camera.

**FACS analysis, islet isolation and gene profiling.** For FACS sorting of GFP<sup>+</sup> cells, pancreata infected by the M3 inducing factors for one month were perfused through the common bile duct, digested with liberase and elastase (Roche), and further dissociated into single cells with EDTA incubation. GFP<sup>+</sup> cells were isolated by FACS with FACSAria (BD Bioscience). Staining of sorted cells indicates that ~70% of total sorted cells are GFP<sup>+</sup> and ~22% are insulin<sup>+</sup>.

Islets were isolated by liberase digestion of the pancreas of Pdx1-GFP animals. Islets were picked manually under a fluorescent dissecting scope. Pancreatic cells devoid of GFP<sup>+</sup> islets were collected as the non-islet sample.

RNA was extracted with Trizol reagent (Invitrogen). Biotin-labelled cRNA probes were synthesized with the Illumina TotalPrep RNA amplification kit (Ambion). Gene profiling was performed with Sentrix BeadChip Array MouseRef-8 v1.1 (Illumina) that contains probes for ~19,000 genes. Data were analysed with BeadStudio software. For identifying differentially enriched genes, the following parameters suggested by Illumina were used: *P* value < 0.05, Diff score > 30, average signal > 100.

**RT-PCR.** Pancreatic tissues were harvested and immediately frozen in liquid nitrogen. Total RNA was extracted with the RNeasy kit (Qiagen). First-strand cDNA was synthesized with Superscript III kit (Invitrogen). Thirty cycles of semiquantitative RT-PCR were performed using the standard protocol. The following primer pairs were used: *Ngn3* viral transgene: ~350 bp, *Ngn3*:F: CAGACGCTGCGCATAGCGGACCAC, IRES2.R: GCGGCTTCGGCCAGTAA CGTTAG. *Pdx1* viral transgene: ~1.2 kb, *Pdx1*:F: GGAGCAAGATT GTGCGGTGACCTC, IRES2.R: GCGGCTTCGGCCAGTAAACGTTAG. *Mafa* viral transgene: ~300 bp, *Mafa*:F: ACATTCTGGAGAGCGAGAAGTGCC, IRES2.R: GCGGCTTCGGCCAGTAAACGTTAG. *GADPH*: ~400 bp, F: ACCA CAGTCCATGCCATCAC, R: TCCACCACCTGTTGCTGTA.

# Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT

Andrew J. Gillis<sup>1</sup>, Anthony P. Schuller<sup>1</sup> & Emmanuel Skordalakes<sup>1</sup>

**A common hallmark of human cancers is the overexpression of telomerase, a ribonucleoprotein complex that is responsible for maintaining the length and integrity of chromosome ends. Telomere length deregulation and telomerase activation is an early, and perhaps necessary, step in cancer cell evolution. Here we present the high-resolution structure of the *Tribolium castaneum* catalytic subunit of telomerase, TERT. The protein consists of three highly conserved domains, organized into a ring-like structure that shares common features with retroviral reverse transcriptases, viral RNA polymerases and B-family DNA polymerases. Domain organization places motifs implicated in substrate binding and catalysis in the interior of the ring, which can accommodate seven to eight bases of double-stranded nucleic acid. Modelling of an RNA–DNA heteroduplex in the interior of this ring demonstrates a perfect fit between the protein and the nucleic acid substrate, and positions the 3′-end of the DNA primer at the active site of the enzyme, providing evidence for the formation of an active telomerase elongation complex.**

Telomerase is active in the early stages of life to maintain telomere length and therefore the chromosomal integrity of frequently dividing cells, and it becomes dormant in most somatic cells during adulthood<sup>1,2</sup>. The ability of telomeres to provide genomic stability is diminished over time owing to both the natural loss of telomeric structure with every cell division, and the loss of telomerase activity—a process which leads to ageing<sup>3,4</sup>. In cancer cells, however, telomerase becomes reactivated and works tirelessly to maintain the short length of telomeres of rapidly dividing cells, leading to their immortality<sup>5,6</sup>. The essential role of telomerase in cancer and ageing makes it an important target for the development of therapies to treat cancer and other age-associated disorders.

Telomerase functions as both a monomer and a dimer<sup>7–10</sup>, and consists of a protein subunit (TERT) and an integral RNA component (TER) which contains the template that TERT uses to add several DNA repeats to the 3′-end of linear chromosomes<sup>11,12</sup>. TERT, the catalytic subunit of telomerase, is highly conserved among phylogenetic groups and shares common motifs with conventional reverse transcriptases, suggesting an overall conservation of the basic catalytic mechanism between these two classes of enzymes<sup>13,14</sup>. Although TER varies considerably in size, sequence and structure between species, core structural elements are conserved, suggesting that there is a common mechanism of telomere replication among organisms<sup>15,16</sup>.

A functional telomerase holoenzyme requires the stable association of the ribonucleoprotein complex, a process mostly carried out by the RNA-binding domain (TRBD)<sup>17,18</sup>. Weak interactions have been reported between TER and both the far amino-terminal domain (a low conservation region of TERT) and the polymerase domain (reverse transcriptase)<sup>18,19</sup>. Current evidence suggests that TRBD binds to the template boundary element of TER, usually a stem loop or a pseudoknot flanked by regions of single-stranded RNA<sup>20–23</sup>. The TRBD–TER association also promotes repeat addition processivity, which is a unique feature of telomerase<sup>19–22,24</sup>. Telomerase repeat addition processivity is also attributed to the IFD (insertion in fingers domain) motif of reverse transcriptase and the carboxy-terminal extension (CTE) proposed to constitute the putative ‘thumb’ domain of telomerase<sup>25–27</sup>.

Initiation of telomere synthesis requires the loading of telomerase onto the end of the chromosomes and the pairing of the 3′-end of the linear DNA substrate with the templating region (usually one and a half repeats of the telomeric repeat)<sup>28–30</sup>. Pairing of the DNA with the RNA template places the 3′-end of the DNA substrate at the active site of the enzyme for nucleotide addition, whereas the RNA template provides the platform for the successive rounds of nucleotide addition and selectivity. RNA–DNA pairing alone is not sufficient for a stable and active telomerase elongation complex and requires extensive contacts of the DNA substrate with both the reverse transcriptase and the putative thumb domain of TERT<sup>25,31</sup>. In some organisms, contacts between the far N-terminal domain and a DNA site upstream of the RNA–DNA hybridization region allow the enzyme to remain attached to the end of the chromosomes during translocation<sup>32,33</sup>.

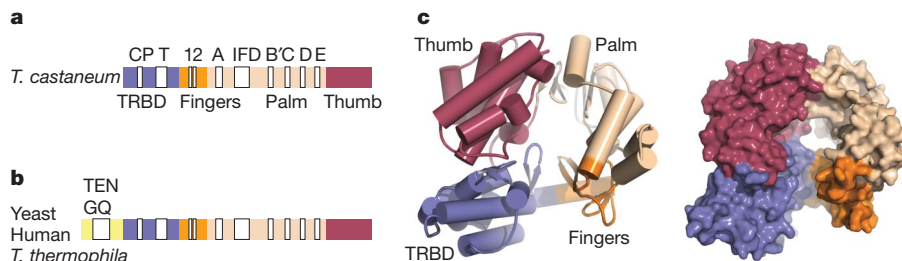
Here we present, to our knowledge, the first high-resolution structure of the catalytic subunit of telomerase. This structure, together with previous biochemical data, provides insights into TERT–TER–DNA assembly and elongation complex formation.

## Architecture of the TERT structure

We have solved the structure of the full-length catalytic subunit of the *T. castaneum* active telomerase<sup>34,35</sup>, TERT, to 2.71 Å resolution. There is a dimer in the asymmetric unit; however, the protein alone is clearly monomeric in solution as indicated by gel filtration and dynamic light scattering (results not shown) suggesting that the dimer we observe in the crystal is the result of crystal packing. This notion is further supported by the fact that a different crystal form (Supplementary Table 1) of the same protein also contains a dimer in the asymmetric unit of a different configuration than the one presented here. It is worth noting that the TERT from this organism does not contain an N-terminal domain, a low conservation region of telomerase (Fig. 1a, b).

The TERT structure is composed of three distinct domains: an RNA-binding domain (TRBD), the reverse transcriptase domain and the CTE thought to represent the putative thumb domain of TERT (Fig. 1a, c). The TRBD is mostly helical and contains an indentation on its surface formed by two conserved motifs (CP

<sup>1</sup>Gene Expression and Regulation Program, The Wistar Institute, 3601 Spruce Street, Philadelphia, Pennsylvania 19104, USA.



**Figure 1 | The structure of TERT.** **a**, The primary structure and conserved motifs of the *T. castaneum* TERT are shown. **b**, The primary structure of human, yeast and *T. thermophila* is shown for comparison. **c**, TERT domain

organization (cartoon and surface representation); the RNA-binding domain (TRBD) is shown in violet, the fingers domain is in orange, the palm domain is in tan and the thumb domain is in red.

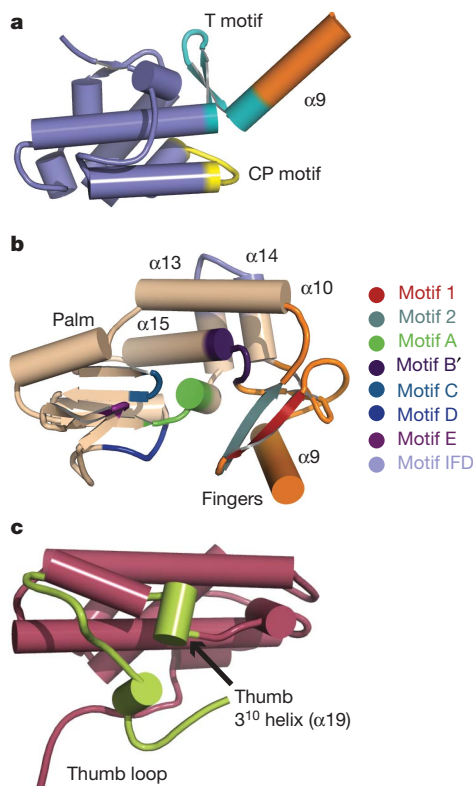
and T) known to bind the double- and single-stranded RNA regions of the template boundary element, respectively<sup>24</sup> (Fig. 2a). Structural comparison of the TRBD from *T. castaneum* with that of the previously determined structure from *Tetrahymena thermophila*<sup>24</sup> shows similarity between the two structures (root mean squared deviation (r.m.s.d.) 2.7 Å), suggesting that a high degree of structural conservation occurs between these domains across organisms of diverse phylogenetic groups.

The reverse transcriptase domain is a mixture of  $\alpha$ -helices and  $\beta$ -strands organized into two subdomains that are most similar to the 'fingers' and 'palm' subdomains of retroviral reverse transcriptases<sup>36</sup>, viral RNA polymerases<sup>37</sup> and B-family DNA polymerases<sup>38</sup>

(Supplementary Fig. 1a–d), and contains important signature motifs that are hallmarks of these families of proteins<sup>14</sup> (Fig. 2b). Structural comparisons of TERT with the HIV reverse transcriptases show that the fingers subdomain of TERT is arranged in the open configuration with respect to the palm subdomain, which is in good agreement with the conformation adopted by HIV reverse transcriptases in the absence of bound nucleotide and nucleic acid substrates<sup>39</sup>. One notable difference between the putative palm domain of TERT and the HIV reverse transcriptases is a long insertion between motifs A and B' of TERT; this is referred to as the IFD motif and is required for telomerase processivity<sup>27</sup>. In the TERT structure, the IFD insertion consists of two antiparallel  $\alpha$ -helices ( $\alpha$ 13 and  $\alpha$ 14) located on the outside periphery of the ring and at the interface of the fingers and the palm subdomains (Fig. 2b). These two helices are almost in a parallel position with the central axis of the plane of the ring, make extensive contacts with helices  $\alpha$ 10 and  $\alpha$ 15, and have an important role in the structural organization of this part of the reverse transcriptase domain. A similar structural arrangement is also present in viral polymerases, and the equivalent of helix  $\alpha$ 10 in these structures is involved in direct contacts with the nucleic acid substrate<sup>40</sup> (Supplementary Fig. 1c).

In contrast to the reverse transcriptase domain, the CTE is an elongated helical bundle that contains several surface-exposed long loops (Fig. 2c). A search in the protein structure database using the secondary-structure matching software (<http://www.ebi.ac.uk/msd-srv/ssm>)<sup>41</sup> produced no structural homologues, suggesting that the CTE domain of telomerase adopts a new fold. Structural comparison of TERT with the HIV reverse transcriptase, with the viral RNA polymerases and with the B-family DNA polymerases places the thumb domain of these enzymes and the CTE domain of TERT in the same spatial position with respect to the fingers and palm subdomains. This suggests that the CTE domain of telomerase is the thumb domain of the enzyme, a finding that is in good agreement with previous biochemical studies<sup>25</sup> (Supplementary Fig. 2).

TERT domain organization brings the TRBD and thumb domain—which constitute the terminal domains of the molecule—together, an arrangement that leads to the formation of a ring-like structure that is reminiscent of the shape of a doughnut (Fig. 1a, b). Several lines of evidence suggest that the domain organization of the TERT structure presented here is biologically relevant. First, the domains of four TERT monomers observed in two different crystal forms (two in each asymmetric unit) all have the same organization (average r.m.s.d. = 0.76 Å between all four monomers). Second, contacts between the N- and the C-terminal domains of TERT are extensive (1,677 Å<sup>2</sup>) and largely hydrophobic in nature, an observation that is consistent with previous biochemical studies<sup>42</sup> (Supplementary Fig. 3). Third, TERT domain organization is similar to that of the polymerase domain (p66 minus the RNase H domain) of its closest homologue, HIV reverse transcriptase<sup>36</sup>. It is also similar to the domain organization of the viral RNA polymerases<sup>37</sup> and that of the B-family DNA polymerases, particularly RB69 (ref. 38; Supplementary Fig. 1a–d). The arrangement of the TERT domains creates a hole in the interior of the particle that is ~26 Å wide and



**Figure 2 | The TERT domain fold and main signature motifs.** **a**, The almost-all helical TRBD binds RNA using two conserved motifs: T (shown in cyan) and CP (shown in yellow). **b**, The reverse transcriptase domain consists of the palm and fingers subdomains. The fingers subdomain contains motifs 1 (red) and 2 (grey) and is implicated in nucleotide and RNA binding. The palm subdomain contains motifs A (green), B' (dark purple), C (blue), D (navy blue), E (magenta) and IFD (light blue) and is involved in nucleotide and nucleic acid binding and DNA synthesis. **c**, The helical thumb domain of TERT contains a loop (thumb loop; light green) which is involved in domain organization and DNA binding.



~21 Å deep, sufficient to accommodate double-stranded nucleic acid approximately seven to eight bases long and in good agreement with existing biochemical data<sup>43,44</sup>.

### The TERT ring binds double-stranded nucleic acid

To understand better how the TERT ring associates with RNA–DNA to form a functional elongation complex, we modelled double-stranded nucleic acid into its interior using the complex of HIV reverse transcriptase with DNA<sup>36</sup>, the closest structural homologue of TERT (Fig. 3a). The TERT–RNA–DNA model immediately shows some notable features that support our model of TERT–nucleic-acid associations. The hole of the TERT ring, and where the nucleic acid heteroduplex is projected to bind, is lined with several key signature motifs that are hallmarks of this family of polymerases and have been implicated in nucleic acid association, nucleotide binding and DNA synthesis (Fig. 3a). Moreover, the organization of these motifs results in the formation of a spiral in the interior of the ring that resembles the geometry of the backbone of double-stranded nucleic acid (Fig. 3b). Several of the motifs, identified as contact points with the DNA substrate, are formed partly by positively charged residues, the side chains of which extend towards the centre of the ring and are poised for direct contact with the backbone of the DNA substrate. For example, the side chain of the highly conserved K210 (Supplementary Fig. 4) that forms part of helix  $\alpha$ 10, is within coordinating distance of the backbone of the modelled DNA, thus providing the stability required for a functional telomerase enzyme. Helix  $\alpha$ 10 lies in the upper segment of the reverse transcriptase domain and faces the interior of the ring. The location and stabilization of this helix is heavily influenced by its extensive contacts with the IFD motif implicated in telomerase processivity<sup>27</sup>. Disruption of the IFD contacts with helix  $\alpha$ 10—by deletion or mutation of this motif—would lead to displacement of helix  $\alpha$ 10 from its current location, which would in turn effect DNA-binding and telomerase function.

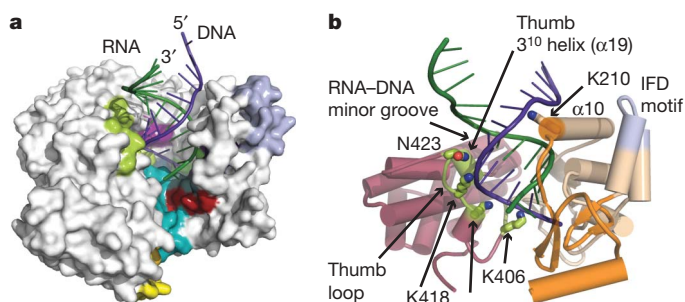
Structural elements of the thumb domain that localize to the interior of the ring also make several contacts with the modelled DNA substrate (Fig. 3a). In particular, the loop (thumb loop) that connects the palm to the thumb domain and constitutes an extension of motif E, also known as the ‘primer grip’ region of telomerase, preserves the geometry of the backbone of double-stranded nucleic acid to a notable degree (Fig. 3b). The side chains of several lysines and asparagines that form part of this loop extend towards the centre of the TERT molecule and are in coordinating distance of the backbone of modelled double-stranded nucleic acid. Of particular interest is K406, located in proximity of motif E. The side chain of this lysine extends towards the

nucleic acid heteroduplex and it is poised for direct contacts with the backbone of the nucleotides located at the 3′-end of the incoming DNA primer. It is therefore possible that the side chain of this lysine, together with motif E, help facilitate placement of the 3′-end of the incoming DNA substrate at the active site of the enzyme during telomere elongation. Sequence alignments of the thumb domain of TERTs from a wide spectrum of phylogenetic groups show that the residues predicted to contact the DNA substrate are always polar (Supplementary Fig. 4). Another interesting feature of the thumb domain, which supports double-stranded nucleic acid binding, is helix  $\alpha$ 19 (Fig. 2c). This is a 3<sup>10</sup> helix (thumb 3<sup>10</sup> helix) that extends into the interior of the ring and seems to dock itself into the minor groove of the modelled double-stranded nucleic acid, thus facilitating RNA–DNA hybrid binding and stabilization (Fig. 3b). Deletion or mutation of the corresponding residues in both yeast and human TERT results in severe loss of TERT processivity, clearly indicating the important role of this motif in TERT function<sup>25,26,45</sup>.

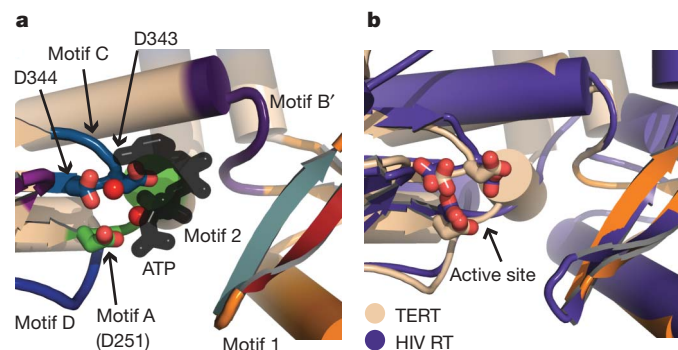
### The active site of TERT and nucleotide binding

The TERT structure presented here was crystallized in the absence of nucleotide substrates and magnesium; however, the location and organization of TERT’s active site and nucleotide-binding pocket can be predicted on the basis of existing biochemical data<sup>14</sup> and structural comparison with the polymerase domain of its closest homologue, the HIV reverse transcriptase<sup>46</sup>. The TERT active site consists of three invariant aspartic acids (D251, D343 and D344) that form part of motifs A and C, which are two short loops located on the palm subdomain and adjacent to the fingers of TERT (Fig. 4a). Structural comparisons of TERT with HIV reverse transcriptases, as well as with RNA and DNA polymerases, show a high degree of similarity between the active sites of these families of proteins (Fig. 4b), suggesting that telomerase also uses a two-metal mechanism for catalysis. Alanine mutants of these TERT aspartic acids resulted in complete loss of TERT activity, indicating that the role of these residues in telomerase function is essential<sup>14</sup>.

The telomerase nucleotide-binding pocket is located at the interface of the fingers and palm subdomains of TERT (Fig. 4a) and consists of conserved residues that form the motifs 1, 2, A, C, B′ and D which are implicated in template and nucleotide binding<sup>47,48</sup> (Supplementary Fig. 5). Structural comparisons of TERT with viral HIV reverse transcriptases bound to ATP<sup>46</sup> support the presence of a nucleotide substrate in this location. Two highly conserved surface-exposed residues Y256 and V342 of motifs A and C, respectively, form a hydrophobic pocket adjacent to and above the three catalytic



**Figure 3 | Model of the TERT–RNA–DNA complex.** **a**, Surface representation of TERT with the modelled RNA (dark green) and DNA (dark purple) shown in cartoon. The main signature motifs are coloured as in Fig. 2. The RNA-binding pocket of TERT is located in the deep cavity on the side of the ring. **b**, Contacts between the TERT ring and a modelled RNA–DNA heteroduplex are shown. The path of the incoming DNA primer within the TERT ring is lined up with several conserved positively charged residues poised for direct contacts with its backbone. The thumb loop and the thumb 3<sup>10</sup> helix are shown in light green; the IFD motif is shown in light blue.



**Figure 4 | The active site and nucleotide-binding pocket of telomerase.** **a**, The active site of telomerase is formed by three invariant aspartic acids that form part of motifs A (D251) and C (D343 and D344). The nucleotide-binding pocket is located at the interface of the fingers and palm subdomains of reverse transcriptase; a modelled nucleotide using the HIV reverse transcriptase (PDB code 2IAJ) is shown as a black stick. **b**, The overlay of the active site residues of TERT (tan) and HIV reverse transcriptase (RT; purple; PDB code 1N5Y) shows that there is a high degree of similarity between the pockets of the two enzymes.

aspartates and this could accommodate the base of the nucleotide substrate. Binding of the nucleotide in this oily pocket places the triphosphate moiety in proximity of the enzyme's active site for coordination with one of the  $Mg^{2+}$  ions. In contrast, it positions the ribose group within coordinating distance of an invariant glutamine (Q308) that forms part of motif B', which is thought to be an important determinant of substrate specificity<sup>49</sup>. Protein contacts with the triphosphate moiety of the nucleotide are mediated by motif D, a long loop located beneath the active site of the enzyme. In particular, the side chain of the invariant K372 is within coordinating distance of the  $\gamma$ -phosphate of the nucleotide, an interaction that probably helps position and stabilize the triphosphate group during catalysis. The side chains of the highly conserved K189 and R192 of motifs 1 and 2, which together form a long  $\beta$ -hairpin that forms part of the fingers subdomain, are also within coordinating distance of both the sugar and triphosphate moieties of the modelled nucleotide. Contacts with either or both the sugar moiety and the triphosphate moiety of the nucleotide substrate would facilitate nucleotide binding and positioning for coordination to the 3'-end of the incoming DNA primer.

#### TRBD facilitates template positioning at the active site

As with most DNA and RNA polymerases, nucleic acid synthesis by telomerase requires pairing of the templating region (usually seven to eight bases or more) of TER with the incoming DNA primer<sup>28</sup>. TRBD–reverse-transcriptase domain organization forms a deep cavity on the surface of the protein that spans the entire width of the wall of the molecule, forming a gap that allows entry into the hole of the ring from its side (Fig. 3a). The arrangement of this cavity with respect to the central hole of the ring provides an elegant mechanism upon TERT–TER assembly for the placement of the RNA template in the interior of the ring and where the enzyme's active site is located. Of particular significance is the arrangement of the  $\beta$ -hairpin that forms part of the T motif. This hairpin extends from the RNA-binding pocket and makes extensive contacts with the thumb loop and motifs 1 and 2 (Fig. 3a). Contacts between this hairpin and both the fingers and the thumb domains place the opening of the TRBD pocket that faces the interior of the ring in proximity to the active site of the enzyme (Fig. 5). It is therefore possible that this  $\beta$ -hairpin acts as an allosteric effector switch that couples RNA binding in the interior of the ring and placement of the RNA template at the active site of the enzyme. Placement of the template into the interior of the molecule would facilitate its pairing with the incoming DNA substrate, which together would form the RNA–DNA hybrid required for telomere elongation. RNA–DNA pairing is a prerequisite of telomere synthesis in that it brings the 3'-end of the incoming DNA primer in proximity to the active site of the enzyme for nucleotide

addition, and the RNA component of the heteroduplex provides the template for the addition of identical repeats of DNA at the ends of chromosomes. Notably, modelling of the RNA–DNA heteroduplex in the interior of the TERT ring places the 5'-end of the RNA substrate at the entry of the RNA-binding pocket and where TERT is expected to associate with TER, whereas it places the 3'-end of the incoming DNA primer at the active site of TERT providing a snapshot of the organization of a functional telomerase elongation complex (Fig. 5).

#### Conclusions

The structure presented here provides a view of the full-length catalytic subunit of telomerase. The structure shows that TERT is organized into an unexpected ring configuration that resembles—both structurally and functionally—the HIV reverse transcriptases, the viral RNA polymerases and the B-family DNA polymerases, suggesting that there is an evolutionary link between these families of enzymes. It also provides insights into the mechanism of TERT and RNA–DNA association, which in turn explains how TERT may assemble with RNA–DNA and offers a snapshot of a functional telomerase elongation complex required for telomere synthesis. Moreover, because telomerase has a critical role in both cancer and ageing, these findings could potentially assist our efforts to identify and develop inhibitors and/or activators of this enzyme for the treatment of cancer and ageing, respectively.

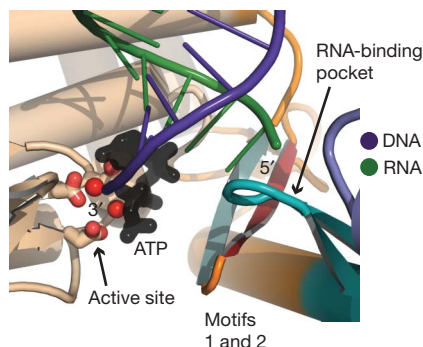
#### METHODS SUMMARY

The full-length TERT of *T. castaneum* was overexpressed in bacteria and purified by nickel, ion-exchange and gel-filtration chromatography. Co-crystallization of the protein–telomeric-DNA ((TCAGG)<sub>3</sub>) produced two crystal forms (orthorhombic and hexagonal), which were grown by the vapour diffusion, sitting-drop method. Data were collected at the National Synchrotron Light Source (NSLS) at beamline X6A and were processed with MOSFILM (Supplementary Table 1). Phases for the orthorhombic crystal were obtained by the method of single isomorphous replacement with anomalous signal using a mercury derivative (CH<sub>3</sub>HgCl; Supplementary Table 1). The model from the orthorhombic crystal was subsequently used to solve the hexagonal crystal form by molecular replacement. Both models were refined to good stereochemistry (Supplementary Table 1).

Full Methods and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 3 June 2008; accepted 23 July 2008.

Published online 31 August 2008.



**Figure 5 | Localization of the RNA–DNA ends in the interior of the ring.** TERT domain organization places the RNA-binding pocket (cyan) and the active site of the enzyme in close proximity to each other. Modelling of the RNA–DNA heteroduplex in the TERT ring places the 3'-end of the DNA substrate (dark purple) at the active site of the enzyme, and the 5'-end of the RNA substrate (dark green) at the RNA-binding pocket of telomerase.

- Blackburn, E. H. Telomeres: no end in sight. *Cell* **77**, 621–623 (1994).
- Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405–413 (1985).
- Wright, W. E. & Hayflick, L. Nuclear control of cellular aging demonstrated by hybridization of anucleate and whole cultured normal human fibroblasts. *Exp. Cell Res.* **96**, 113–121 (1975).
- Wright, W. E. & Shay, J. W. Cellular senescence as a tumor-protection mechanism: the essential role of counting. *Curr. Opin. Genet. Dev.* **11**, 98–103 (2001).
- Bodnar, A. G. *et al.* Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349–352 (1998).
- Campisi, J., Kim, S. H., Lim, C. S. & Rubio, M. Cellular senescence, cancer and aging: the telomere connection. *Exp. Gerontol.* **36**, 1619–1637 (2001).
- Beattie, T. L., Zhou, W., Robinson, M. O. & Harrington, L. Functional multimerization of the human telomerase reverse transcriptase. *Mol. Cell. Biol.* **21**, 6151–6160 (2001).
- Bryan, T. M., Goodrich, K. J. & Cech, T. R. *Tetrahymena* telomerase is active as a monomer. *Mol. Biol. Cell* **14**, 4794–4804 (2003).
- Moriarty, T. J., Huard, S., Dupuis, S. & Autexier, C. Functional multimerization of human telomerase requires an RNA interaction domain in the N terminus of the catalytic subunit. *Mol. Cell. Biol.* **22**, 1253–1265 (2002).
- Prescott, J. & Blackburn, E. H. Functionally interacting telomerase RNAs in the yeast telomerase complex. *Genes Dev.* **11**, 2790–2800 (1997).
- Feng, J. *et al.* The RNA component of human telomerase. *Science* **269**, 1236–1241 (1995).
- Nakamura, T. M. *et al.* Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277**, 955–959 (1997).
- Counter, C. M., Meyerson, M., Eaton, E. N. & Weinberg, R. A. The catalytic subunit of yeast telomerase. *Proc. Natl Acad. Sci. USA* **94**, 9202–9207 (1997).

14. Lingner, J. *et al.* Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276**, 561–567 (1997).
15. Chen, J. L. & Greider, C. W. An emerging consensus for telomerase RNA structure. *Proc. Natl Acad. Sci. USA* **101**, 14683–14684 (2004).
16. Ly, H., Blackburn, E. H. & Parslow, T. G. Comprehensive structure–function analysis of the core domain of human telomerase RNA. *Mol. Cell. Biol.* **23**, 6849–6856 (2003).
17. Lai, C. K., Mitchell, J. R. & Collins, K. RNA binding domain of telomerase reverse transcriptase. *Mol. Cell. Biol.* **21**, 990–1000 (2001).
18. O'Connor, C. M., Lai, C. K. & Collins, K. Two purified domains of telomerase reverse transcriptase reconstitute sequence-specific interactions with RNA. *J. Biol. Chem.* **280**, 17533–17539 (2005).
19. Friedman, K. L. & Cech, T. R. Essential functions of amino-terminal domains in the yeast telomerase catalytic subunit revealed by selection for viable mutants. *Genes Dev.* **13**, 2863–2874 (1999).
20. Chen, J. L. & Greider, C. W. Template boundary definition in mammalian telomerase. *Genes Dev.* **17**, 2747–2752 (2003).
21. Lai, C. K., Miller, M. C. & Collins, K. Template boundary definition in *Tetrahymena* telomerase. *Genes Dev.* **16**, 415–420 (2002).
22. Seto, A. G. *et al.* A template-proximal RNA paired element contributes to *Saccharomyces cerevisiae* telomerase activity. *RNA* **9**, 1323–1332 (2003).
23. Tzfati, Y., Fulton, T. B., Roy, J. & Blackburn, E. H. Template boundary in a yeast telomerase specified by RNA structure. *Science* **288**, 863–867 (2000).
24. Rouda, S. & Skordalakes, E. Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. *Structure* **15**, 1403–1412 (2007).
25. Hossain, S., Singh, S. & Lue, N. F. Functional analysis of the C-terminal extension of telomerase reverse transcriptase. A putative “thumb” domain. *J. Biol. Chem.* **277**, 36174–36180 (2002).
26. Huard, S., Moriarty, T. J. & Autexier, C. The C terminus of the human telomerase reverse transcriptase is a determinant of enzyme processivity. *Nucleic Acids Res.* **31**, 4059–4070 (2003).
27. Lue, N. F., Lin, Y. C. & Mian, I. S. A conserved telomerase motif within the catalytic domain of telomerase reverse transcriptase is specifically required for repeat addition processivity. *Mol. Cell. Biol.* **23**, 8440–8449 (2003).
28. Lee, M. S. & Blackburn, E. H. Sequence-specific DNA primer effects on telomerase polymerization activity. *Mol. Cell. Biol.* **13**, 6586–6599 (1993).
29. Lingner, J., Hendrick, L. L. & Cech, T. R. Telomerase RNAs of different ciliates have a common secondary structure and a permuted template. *Genes Dev.* **8**, 1984–1998 (1994).
30. Shippen-Lentz, D. & Blackburn, E. H. Functional evidence for an RNA template in telomerase. *Science* **247**, 546–552 (1990).
31. Finger, S. N. & Bryan, T. M. Multiple DNA-binding sites in *Tetrahymena* telomerase. *Nucleic Acids Res.* **36**, 1260–1272 (2008).
32. Jacobs, S. A., Podell, E. R. & Cech, T. R. Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. *Nature Struct. Mol. Biol.* **13**, 218–225 (2006).
33. Lue, N. F. A physical and functional constituent of telomerase anchor site. *J. Biol. Chem.* **280**, 26586–26591 (2005).
34. Osanai, M., Kojima, K. K., Futahashi, R., Yaguchi, S. & Fujiwara, H. Identification and characterization of the telomerase reverse transcriptase of *Bombyx mori* (silkworm) and *Tribolium castaneum* (flour beetle). *Gene* **376**, 281–289 (2006).
35. Richards, S. *et al.* The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**, 949–955 (2008).
36. Sarafianos, S. G. *et al.* Structures of HIV-1 reverse transcriptase with pre- and post-translocation AZTMP-terminated DNA. *EMBO J.* **21**, 6614–6624 (2002).
37. Di Marco, S. *et al.* Interdomain communication in hepatitis C virus polymerase abolished by small molecule inhibitors bound to a novel allosteric site. *J. Biol. Chem.* **280**, 29765–29770 (2005).
38. Wang, J. *et al.* Crystal structure of a pol  $\alpha$  family replication DNA polymerase from bacteriophage RB69. *Cell* **89**, 1087–1099 (1997).
39. Ding, J. *et al.* Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.* **284**, 1095–1111 (1998).
40. Ferrer-Orta, C. *et al.* Structure of foot-and-mouth disease virus RNA-dependent RNA polymerase and its complex with a template-primer RNA. *J. Biol. Chem.* **279**, 47212–47221 (2004).
41. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst. D* **60**, 2256–2268 (2004).
42. Arai, K. *et al.* Two independent regions of human telomerase reverse transcriptase are important for its oligomerization and telomerase activity. *J. Biol. Chem.* **277**, 8538–8544 (2002).
43. Forstemann, K. & Lingner, J. Telomerase limits the extent of base pairing between template RNA and telomeric DNA. *EMBO Rep.* **6**, 361–366 (2005).
44. Hammond, P. W. & Cech, T. R. Euplotes telomerase: evidence for limited base-pairing during primer elongation and dGTP as an effector of translocation. *Biochemistry* **37**, 5162–5172 (1998).
45. Banik, S. S. *et al.* C-terminal regions of the human telomerase catalytic subunit essential for *in vivo* enzyme activity. *Mol. Cell. Biol.* **22**, 6234–6246 (2002).
46. Das, K. *et al.* Crystal structures of clinically relevant Lys103Asn/Tyr181Cys double mutant HIV-1 reverse transcriptase in complexes with ATP and non-nucleoside inhibitor HBV 097. *J. Mol. Biol.* **365**, 77–89 (2007).
47. Bosoy, D. & Lue, N. F. Functional analysis of conserved residues in the putative “finger” domain of telomerase reverse transcriptase. *J. Biol. Chem.* **276**, 46305–46312 (2001).
48. Haering, C. H., Nakamura, T. M., Baumann, P. & Cech, T. R. Analysis of telomerase catalytic subunit mutants *in vivo* and *in vitro* in *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA* **97**, 6367–6372 (2000).
49. Smith, R. A., Anderson, D. J. & Preston, B. D. Hypersusceptibility to substrate analogs conferred by mutations in human immunodeficiency virus type 1 reverse transcriptase. *J. Virol.* **80**, 7169–7178 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Marmorstein for critical reading of this manuscript. This project was funded by the Pennsylvania Department of Health and The Ellison Medical Foundation.

**Author Contributions** E.S. designed the experimental plan, carried out the research and wrote the paper. A.J.G. and A.P.S. assisted with the experimental work.

**Author Information** The atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 3DU5 and 3DU6. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to E.S. ([skorda@wistar.org](mailto:skorda@wistar.org)).



## METHODS

**Protein expression and purification.** The synthetic gene of *T. castaneum* full-length TERT was cloned into a modified version of the pET28b vector containing a cleavable hexahistidine tag at its N terminus. The protein was overexpressed in *Escherichia coli* BL21 (pLysS) at 30 °C for 4 h. The cells were lysed by sonication in 50 mM Tris-HCl, pH 7.5, 10% glycerol, 0.5 M KCl, 5 mM  $\beta$ -mercaptoethanol and 1 mM phenylmethyl sulphonyl fluoride, on ice. The protein was purified over a Ni-NTA column followed by TEV cleavage of the hexahistidine tag overnight at 4 °C. The TERT-TEV mixture was dialysed to remove the excess imidazole and the protein was further purified over a second Ni-NTA column that was used to remove all His-tagged products. The Ni-NTA flow through was then passed over a POROS-HS column (Perseptive Biosystems) to remove any trace amounts of protein contaminants. At this stage the protein was more than 99% pure. The protein was finally purified over a sephedex-S200 sizing column pre-equilibrated with 50 mM Tris-HCl, pH 7.5, 10% glycerol, 0.5 M KCl and 1 mM Tris(2-carboxyethyl)phosphine (TCEP) to remove any TERT aggregates, and the protein was concentrated to 10 mg ml<sup>-1</sup> using an Amicon 30K cutoff (Millipore) and stored at 4 °C for subsequent studies. Stock protein was dialysed in 10 mM Tris-HCl, pH 7.5, 200 mM KCl and 1 mM TCEP before crystallization trials.

**Protein crystallization and data collection.** Initial crystal trials of the protein alone did not produce crystals. Co-crystallization of the protein with single-stranded telomeric DNA ((TCAGG)<sub>3</sub>) produced two rod-like crystal forms, one of which belongs to the orthorhombic space group  $P2_12_12_1$  and diffracted to 2.71 Å, and the other belongs to the hexagonal space group  $P6_1$  and diffracted to 3.25 Å resolution. The protein-nucleic-acid mix was prepared before setting crystal trials by mixing one volume of dialysed protein with a 1.2-fold excess of the DNA substrate. Both crystal forms were grown by the vapour diffusion, sitting-drop method by mixing one volume of the protein-DNA mix with one volume of reservoir solution. Orthorhombic crystals were grown in the presence of 50 mM HEPES, pH 7.0, and 1.5 M NaNO<sub>3</sub>, whereas hexagonal crystals were grown in 100 mM Tris, pH 8.0, and 2 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and both crystal forms were grown at room temperature. Orthorhombic crystals were collected into cryoprotectant solution that contained 50 mM HEPES, pH 7.0, 25% glycerol, 1.7 M NaNO<sub>3</sub>, 0.2 M KCl and 1 mM TCEP and were flash frozen in liquid nitrogen. Hexagonal crystals were collected into cryoprotectant solution that contained 100 mM Tris, pH 8.0, 25% glycerol, 2 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.2 M KCl and 1 mM TCEP and were also flash frozen in liquid nitrogen. Data were collected at the NSLS at

beamline X6A and processed with HKL-2000 (ref. 50; Supplementary Table 1). Both crystal forms contain a dimer in the asymmetric unit.

**Structure determination and refinement.** Initial phases for the orthorhombic crystals were obtained using the method of single isomorphous replacement with anomalous signal using two data sets collected from two different mercury-derivatized crystals at two different wavelengths (Hg1, 1.00850 Å; Hg2, 1.00800 Å; Supplementary Table 1). The derivatives were prepared by soaking the crystals with 5 mM methyl mercury chloride (CH<sub>3</sub>HgCl) for 15 min. Initially, twelve heavy atom sites were located using SOLVE<sup>51</sup>, and they were refined and new phases were calculated with MLPHARE<sup>52</sup>. The MLPHARE-improved phases were used to identify the remaining heavy atom sites (22 in total) by calculating an anomalous difference map to a resolution of 3.5 Å. The MLPHARE phases obtained using all the heavy atom sites were then used in DM (density modification package) with two-fold non-crystallographic symmetry and phase extension, using the high-resolution (2.71 Å) data set collected at 1.00800 Å wavelength to calculate starting experimental maps. These maps were of sufficient quality for model building which was carried out in COOT<sup>53</sup>. The electron density map shows clear density for all 596 residues of the protein. Notably, we did not observe density for the nucleic acid substrate in the structure. The model was refined using both CNS-SOLVE<sup>54</sup> and REFMAC5 (ref. 55). The last cycles of refinement were carried out with TLS restraints as implemented in REFMAC5 (Supplementary Table 1). The  $P2_12_12_1$  refined model was used to solve the structure of the TERT crystallized in the  $P6_1$  crystal form (data collected at 0.97980 Å wavelength) by molecular replacement with PHASER<sup>56</sup>.

50. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
51. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374**, 22–37 (2003).
52. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
53. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
54. Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
55. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
56. Pottorff, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D* **59**, 1131–1137 (2003).

## LETTERS

# An 84- $\mu\text{G}$ magnetic field in a galaxy at redshift $z = 0.692$

Arthur M. Wolfe<sup>1</sup>, Regina A. Jorgenson<sup>1</sup>, Timothy Robishaw<sup>2</sup>, Carl Heiles<sup>2</sup> & Jason X. Prochaska<sup>3</sup>

The magnetic field pervading our Galaxy is a crucial constituent of the interstellar medium: it mediates the dynamics of interstellar clouds, the energy density of cosmic rays, and the formation of stars<sup>1</sup>. The field associated with ionized interstellar gas has been determined through observations of pulsars in our Galaxy. Radio-frequency measurements of pulse dispersion and the rotation of the plane of linear polarization, that is, Faraday rotation, yield an average value for the magnetic field of  $B \approx 3 \mu\text{G}$  (ref. 2). The possible detection of Faraday rotation of linearly polarized photons emitted by high-redshift quasars<sup>3</sup> suggests similar magnetic fields are present in foreground galaxies with redshifts  $z > 1$ . As Faraday rotation alone, however, determines neither the magnitude nor the redshift of the magnetic field, the strength of galactic magnetic fields at redshifts  $z > 0$  remains uncertain. Here we report a measurement of a magnetic field of  $B \approx 84 \mu\text{G}$  in a galaxy at  $z = 0.692$ , using the same Zeeman-splitting technique that revealed an average value of  $B = 6 \mu\text{G}$  in the neutral interstellar gas of our Galaxy<sup>4</sup>. This is unexpected, as the leading theory of magnetic field generation, the mean-field dynamo model, predicts large-scale magnetic fields to be weaker in the past rather than stronger<sup>5</sup>.

We detected Zeeman splitting of the  $z = 0.692$ , 21-cm absorption line in the direction of the quasar 3C 286 (refs 6, 7) using the 100-m Robert C. Byrd Green Bank Telescope (GBT) of the National Radio Astronomy Observatory. The absorption arises in a damped Lyman  $\alpha$  ( $\text{Ly}\alpha$ ) system (henceforth denoted DLA-3C286) that is drawn from a population of neutral gas layers widely thought to be the progenitors of modern galaxies<sup>8</sup>. The radio data for DLA-3C286 are summarized in Fig. 1, which shows the line-depth spectra constructed from the measurable quantities used to describe polarized radiation, that is, the Stokes parameters. We show line-depth spectra constructed from the  $I(\nu)$  and  $V(\nu)$  Stokes parameters (where  $\nu$  denotes frequency) near the 839.4-MHz frequency centroid of the redshifted 21-cm absorption line. Figure 1a shows the line-depth spectrum constructed from  $I(\nu)$ . A Gaussian fit to the absorption line in Fig. 1a yields a redshift of  $z = 0.6921526 \pm 0.0000008$ , a central optical depth of  $\tau_0 = 0.095 \pm 0.006$ , and a velocity dispersion of  $\sigma_\nu = 3.75 \pm 0.20 \text{ km s}^{-1}$ , which are in good agreement with previous results<sup>6,7</sup>.

In Fig. 1b, we plot the line-depth spectrum constructed from  $V(\nu)$ , which shows the classic 'S curve' pattern expected for Zeeman splitting. From our least-squares fit to the data, we find that  $B_{\text{los}} = 83.9 \pm 8.8 \mu\text{G}$ , where  $B_{\text{los}}$  is the magnetic field component projected along the line of sight (we note that the direction of  $B_{\text{los}}$  is unknown because the instrumental sense of circular polarization was not calibrated). This magnetic field differs in two respects from the magnetic fields obtained from Zeeman splitting arising in interstellar clouds in the Galaxy. First, the field strength corresponds to the line-of-sight component of the mean field  $\langle B_{\text{los}} \rangle$  averaged over transverse dimensions exceeding 200 pc, as very-long-baseline interferometry

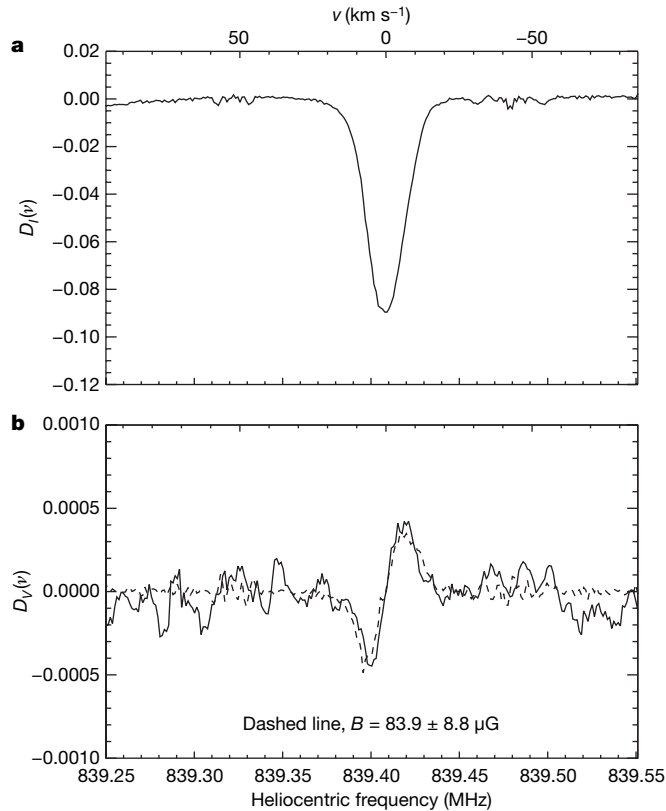
observations of the 21-cm absorption line show that the gas layer must extend across more than  $0.03''$  to explain the difference between the velocity centroids of the fringe amplitude and phase-shift spectra<sup>9</sup> (although the data are consistent with a magnetic field coherence length of less than 200 pc, the resulting gradient in magnetic pressure would produce velocity differences exceeding the shift of  $\sim 3 \text{ km s}^{-1}$  across 200 pc detected by very-long-baseline interferometry). By contrast, the transverse dimensions of radio beams subtended at neutral interstellar clouds in the Galaxy are typically less than 1 pc. Second, this field is at least an order of magnitude stronger than the  $6\text{-}\mu\text{G}$  average of magnetic fields inferred from Zeeman splitting for such clouds<sup>4</sup>.

We obtained further information about conditions in the absorbing gas in DLA-3C286 from accurate optical spectra acquired with the HIRES echelle spectrograph on the Keck I 10-m telescope. Figure 2 shows velocity profiles for several resonance absorption lines arising from dominant low-ionization states of abundant elements. The results of our least-squares fit of Voigt profiles to the data are shown in Table 1, where the optical redshift is displaced  $+3.8 \pm 0.2 \text{ km s}^{-1}$  from the 21-cm redshift. This solution also yields ionic column densities from which we derived the logarithmic metal abundances with respect to solar abundances,  $[\text{M}/\text{H}]$ , and dust-to-gas ratios with respect to the Galactic interstellar medium,  $[\text{D}/\text{G}]$ . These are among the lowest values of  $[\text{M}/\text{H}]$  and  $[\text{D}/\text{G}]$  deduced for damped  $\text{Ly}\alpha$  systems at  $z = 0.7$  (refs 10, 11). The low metallicity indicates a history of low star formation rates. Because the intensity of far-ultraviolet radiation emitted by young massive stars is proportional to the concurrent star formation rate per unit area,  $\Sigma_{\text{SFR}}$ , low values of  $\Sigma_{\text{SFR}}$  should result in low grain photoelectric heating rates per hydrogen atom,  $\Gamma_{\text{pe}}$  (ref. 11). This is consistent with the low upper limit,  $\Gamma_{\text{pe}} < 10^{-27.4} \text{ erg s}^{-1}$  per hydrogen atom, obtained by combining the assumption of thermal balance with the absence of C II\* absorption (that is, absorption from C II in the excited  $^2\text{P}_{3/2}$  fine-structure state) at a wavelength of  $1,335.7 \text{ \AA}$  in the previous low-resolution Hubble Space Telescope spectra of quasar 3C 286 (ref. 12), and indicates that  $\Sigma_{\text{SFR}} < 10^{-2.9} M_\odot \text{ yr}^{-1} \text{ kpc}^{-2}$  (95% confidence level), which is less than the solar-neighbourhood value of  $10^{-2.4} M_\odot \text{ yr}^{-1} \text{ kpc}^{-2}$  (ref. 13).

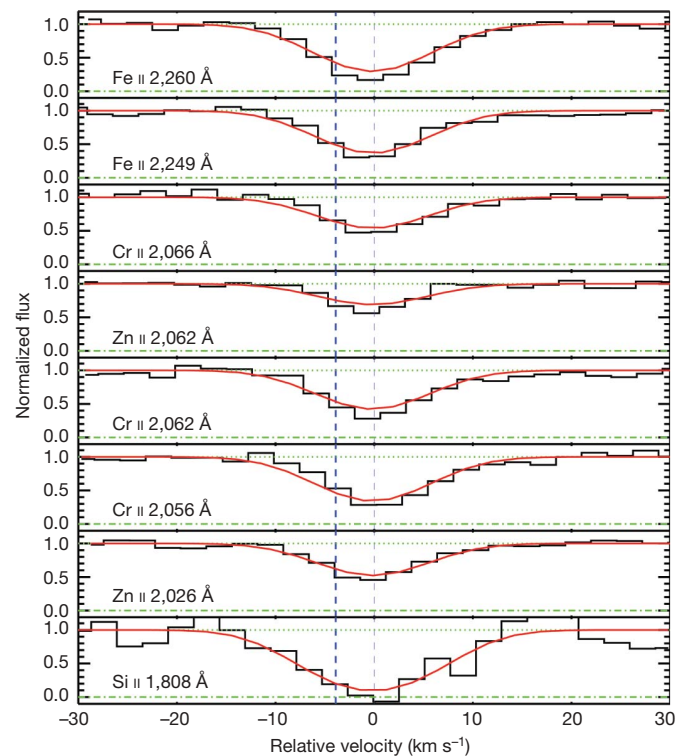
As a result, we have detected an unusually strong magnetic field at  $z = 0.692$  with a coherence length that probably exceeds 200 pc in neutral gas that is quiescent, metal poor, nearly dust free, and presents little evidence of star formation. To model this configuration, we first consider the magnetostatic equilibrium of a plane-parallel sheet with in-plane magnetic field  $B_{\text{plane}}$  orthogonal to the vertical gravitational field exerted by gas with perpendicular mass surface density  $\Sigma$ . In magnetostatic equilibrium, the total mid-plane pressure,  $B_{\text{plane}}^2/8\pi + \rho\sigma_v^2$ , equals the 'weight' of the gas,  $\pi G \Sigma^2/2$ , where  $\rho$  is the mass volume density of the gas and  $G$  is the gravitational constant. However, because the pressure-to-weight ratio exceeds

<sup>1</sup>Department of Physics and Center for Astrophysics and Space Sciences, University of California, San Diego, La Jolla, California 92093-0424, USA. <sup>2</sup>Astronomy Department, University of California, Berkeley, California 94720-3411, USA. <sup>3</sup>UCO-Lick Observatory; University of California, Santa Cruz, Santa Cruz, California 95064, USA.

715 in DLA-3C286, the magnetized gas cannot be confined by its self-gravity. Therefore, self-consistent magnetostatic configurations are ruled out unless the contribution of stars to  $\Sigma$  exceeds  $\sim 350 M_{\odot} \text{ pc}^{-2}$ . Although this is larger than the  $50 M_{\odot} \text{ pc}^{-2}$  surface density perpendicular to the solar neighbourhood, such surface densities are common in the central regions of galaxies. In fact, high surface densities of stars probably confine the highly magnetized gas



**Figure 1 | Line-depth spectra of Stokes parameters.** Data acquired in 12.6 hours of on-source integration with the GBT radio antenna. Because the GBT feeds detect only orthogonal, linearly polarized signals, whereas Zeeman splitting requires measuring circular polarization to construct  $V(v)$ , we generated  $V(v)$  by cross-correlation techniques<sup>23</sup>. The velocity  $v = 0 \text{ km s}^{-1}$  corresponds to  $z = 0.6921526$ . **a**, Line-depth function  $D_V(v) \equiv (I(v) - I_c(v))/I_c(v)$ . Here  $I(v) \equiv s_0 + s_{90}$ , with  $s_0$  the power measured in linear-polarization position angle  $\theta$ , corresponds to the total intensity spectrum, and  $I_c(v)$  is a model fit to the  $I(v)$  continuum.  $D_V(v) = \exp(-\tau(v)) - 1$ , where  $\tau(v) \equiv (\tau(v)_0 + \tau(v)_{90})/2$  is the average optical depth in the two orthogonal states of linear polarization<sup>4</sup>. **b**, Line-depth function  $D_V(v) \equiv V(v)/I_c(v)$ , where  $V(v) \equiv s_{\text{RCP}} - s_{\text{LCP}}$  is the difference in power between the right-hand and left-hand circularly polarized (respectively RCP and LCP) signals. Here  $D_V(v) = -(\tau_V(v)/2)\exp(-\tau(v))$ , where  $\tau_V(v) \equiv \tau_{\text{RCP}}(v) - \tau_{\text{LCP}}(v) \ll 1$  (ref. 4) is the difference between the optical depths of RCP and LCP photons. For Zeeman splitting of the 21-cm line, the degeneracy of the  $F = 0$  to  $F = 1$  hyperfine transition is removed because the  $m_F = -1, 0, +1$  states differ in energy. This results in a small frequency difference between absorbed LCP photons ( $m_F = -1$ ) and RCP photons ( $m_F = +1$ ).  $V(v)$  is crucial for detecting Zeeman splitting because the orthogonal, circularly polarized states of the photon are eigenstates of the spin angular momentum operator with eigenvalues  $\pm \hbar$ , that is, angular momenta directed along or opposite to the direction of photon propagation<sup>24</sup>. When  $B_{105} = B$ , transitions between the hyperfine  $F = 0$  and  $F = 1$  states occur exclusively through absorption of LCP or RCP photons through excitation of the  $m_F = -1$  and  $m_F = +1$  hyperfine states, respectively. Because  $V(v)$  is the difference in the RCP and LCP intensities, the resulting  $V(v)$  line profile is the difference between two Gaussian absorption profiles with frequency centroids shifted by  $\Delta\nu_B = 2.8B_{105}(1+z)^{-1} \text{ Hz}$  (where  $B_{105}$  is measured in microgauss). The ‘S curve’ pattern is due to the sign flip in RCP-minus-LCP intensity difference as  $v$  passes through the line centre.



**Figure 2 | HIRES velocity profiles for dominant low-ionization states of abundant elements in the 21-cm absorber in the direction of quasar 3C 286.** Spectral resolution is  $\Delta\nu = 7.0 \text{ km s}^{-1}$  and the average signal-to-noise ratio per  $2.1\text{-km s}^{-1}$  pixel is about 30:1. The bold dashed vertical line denotes the velocity centroid of the single-dish 21-cm absorption feature and the faint dashed vertical line denotes the velocity centroid of the resonance line shown in the figure. Our least-squares fit of Voigt profiles (red) to the data (black) yields ionic column densities as well as the redshift centroid and velocity dispersion shown in Table 1 (lower and upper green horizontal lines refer to zero and unit normalized fluxes, respectively). Because refractory elements such as Fe and Cr can be depleted onto dust grains<sup>25</sup>, we used the volatile elements Si and Zn to derive a logarithmic metal abundance with respect to solar abundances of  $[M/H] = -1.30$ . The depletion ratios  $[\text{Fe}/\text{Si}]$  and  $[\text{Cr}/\text{Zn}]$  were then used to derive a conservative upper limit on the logarithmic dust-to-gas ratio relative to Galactic values of  $[D/G] < -1.8$ .

in the nuclear rings of barred spirals. These exhibit total field strengths of  $\sim 100 \mu\text{G}$ , inferred by assuming equipartition of magnetic and cosmic-ray energy densities<sup>1</sup>. However, because the rings are associated with regions of active star formation, high molecular content and high dust content, they are unlikely sites for the magnetic field detected in DLA-3C286.

On the other hand, the absorption site might consist of highly magnetized gas confined by the gravity exerted by a disk of old stars. The H I disks found at the centres of early-type S0 and elliptical galaxies<sup>14</sup> are possible prototypes. Support for this idea stems from a high-resolution image obtained with the Hubble Space Telescope: a Wide Field and Planetary Camera 2 (WFPC2) I-band image, from which the quasar has been subtracted, reveals residual emission spread over angular scales of  $\sim 1''$  (ref. 15). The asymmetry of the light distribution with respect to the point-source quasar suggests

**Table 1 | Physical parameters of DLA-3C286 inferred from optical absorption**

Ion, X	$\log_{10}[N(X) \text{ (cm}^{-2}\text{)}]$	$[X/H]$
H I	$21.25 \pm 0.02$	—
Fe II	$15.09 \pm 0.01$	$-1.66 \pm 0.02$
Cr II	$13.44 \pm 0.01$	$-1.48 \pm 0.02$
Zn II	$12.53 \pm 0.03$	$-1.39 \pm 0.03$
Si II	$>15.48$	$>-1.31$

Redshift,  $z = 0.69217485 \pm 0.00000058$ ; velocity dispersion,  $\sigma_v = 3.08 \pm 0.13 \text{ km s}^{-1}$ .  $N(X)$ , column density of ion X.



that some of the light is emitted by a foreground galaxy with a brightness centroid displaced less than  $0.5''$  from the quasar. The location of diffuse emission in the direction of an amorphous object detected  $2.5''$  from the quasar in ground-based imaging<sup>16</sup> further suggests that the diffuse emission comes from central regions of the amorphous object. A recent reanalysis of the WFPC2 image shows the amorphous object to be a filament resembling a spiral arm or tidal tail (H.-W. Chen, personal communication), that is, the outer appendage of a galaxy centred within a few kiloparsecs of the quasar sightline.

However, the magnetic field detected in DLA-3C286 may not be confined by gravity in an equilibrium configuration. Rather, the detected field may be enhanced by a shock (F. H. Shu, personal communication). Assuming a typical value of  $B_{\text{plane}} \approx 5 \mu\text{G}$  for the equilibrium field of the pre-shock gas, we find that a shock-front velocity of  $\sim 250 \text{ km s}^{-1}$  will result in a post-shock field strength of  $\sim 100 \mu\text{G}$  in the limit of flux freezing in a radiative shock with post-shock density of  $\sim 10 \text{ cm}^{-3}$ . This scenario seems plausible because  $250 \text{ km s}^{-1}$  is a reasonable value for the impact velocity generated by the merger between the gaseous disks of two late-type galaxies, and the WFPC2 image is consistent with the presence of two foreground galaxies. But the second disk would create another set of absorption lines displaced  $\geq 250 \text{ km s}^{-1}$  from the redshift of DLA-3C286, which is the only redshift observed. By contrast, the merger between a gaseous disk and an elliptical galaxy could result in only one damped Ly $\alpha$  system redshift, as a significant fraction of elliptical galaxies do not contain H I disks<sup>12</sup>. In this case, a shock front moving in the plane of the disk galaxy would be generated by the gravitational impulse induced by the elliptical galaxy moving normal to the plane. Preliminary estimates indicate that an elliptical galaxy with a modest mass,  $M = 2 \times 10^{11} M_{\odot}$ , and impact velocity of  $\sim 300 \text{ km s}^{-1}$  would produce a cylindrical shock of sufficient strength to boost an initial field with  $B_{\text{plane}} \approx 10 \mu\text{G}$  to a final field of  $\sim 100 \mu\text{G}$ .

Let us examine these scenarios more closely. The quiescent velocity field of the gas fits in naturally with the ‘magnetostatic equilibrium’ scenario, because the low value of  $\Sigma_{\text{SFR}}$  suggests a low rate of energy injection into the gas by supernovae<sup>17</sup>, which could result in a velocity dispersion of  $\sigma_v \approx 4 \text{ km s}^{-1}$ . Moreover, the weak radio jets associated with early-type galaxies containing central H I disks are natural sources of magnetic fields for these disks. However, 21-cm absorption measurements of such disks in nearby galaxies reveal the presence of absorption line widths far broader than the narrow line width of DLA-3C286 (ref. 18). Also, it is unclear whether or not the high surface density of old stars required to confine the magnetic fields are present in these disks, and whether or not the build-up of  $B_{\text{plane}}$  to  $100 \mu\text{G}$  is possible in the 4–5-Gyr age of the disk. In the ‘merger scenario’, the dynamo need only build up to  $\sim 10 \mu\text{G}$  in the same time interval, but it is then necessary to explain why the post-shock velocity field averaged over length scales of  $200 \text{ pc}$  is so quiescent. Furthermore, the probability,  $p$ , of detecting  $\sim 100\text{-}\mu\text{G}$  magnetic fields in a random sample of 21-cm absorbers is small. Our estimates, based on the merger fraction of galaxies with  $z \approx 1$  (ref. 19) and on the duration time for magnetic field enhancement, suggest that  $p \approx 0.005\text{--}0.03$ : either we were lucky, or some characteristic of DLA-3C286, such as narrow line width, is a signature of strong magnetic fields.

Therefore, it is premature to decide among these and other possible models to explain the presence of the  $84\text{-}\mu\text{G}$  magnetic field in DLA-3C286. However, our data support the inference from recent tentative evidence for Faraday rotation in high- $z$  quasars<sup>20</sup> that magnetic fields are generic features of galaxies at high redshifts, which potentially have a more important role in galaxy formation and evolution<sup>21</sup> than hitherto realized. Specifically, the highly magnetized gas that we have detected could suppress gravitational collapse and, hence, may be a reason for the low *in situ* star formation rates

of high- $z$  damped Ly $\alpha$  systems<sup>22</sup>. We plan to test this hypothesis by using the GBT to search for Zeeman splitting in high-redshift damped Ly $\alpha$  systems exhibiting 21-cm absorption.

Received 4 April; accepted 15 July 2008.

- Beck, R. in *Cosmic Magnetic Fields* 41–68 (Lect. Notes Phys. 664, Springer, 2005).
- Han, J. L., Manchester, R. N., Lyne, A. G., Qiao, G. J. & van Straten, W. Pulsar rotation measures and the large-scale structure of galactic magnetic fields. *Astrophys. J.* **642**, 868–881 (2006).
- Kronberg, P. P. et al. A global probe of cosmic magnetic fields to high redshifts. *Astrophys. J.* **676**, 70–79 (2008).
- Heiles, C. & Troland, T. H. The millennium Arecibo 21 centimeter absorption-line survey. III. Techniques for spectral polarization and results for Stokes V. *Astrophys. J. Suppl. Ser.* **151**, 271–297 (2004).
- Parker, E. The origin of magnetic fields. *Astrophys. J.* **160**, 383–404 (1970).
- Brown, R. L. & Roberts, M. S. 21-centimeter absorption at  $z=0.692$  in the quasar 3C 286. *Astrophys. J.* **184**, L7–L10 (1973).
- Davis, M. M. & May, L. S. New observations of the radio absorption line in 3C 286, with potential application to the direct measurement of cosmological deceleration. *Astrophys. J.* **219**, 1–4 (1978).
- Wolfe, A. M., Gawiser, E. & Prochaska, J. X. Damped Ly $\alpha$  systems. *Annu. Rev. Astron. Astrophys.* **43**, 861–918 (2005).
- Wolfe, A. M., Broderick, J. J., Condon, J. J. & Johnston, K. J. 3C 286: A cosmological QSO? *Astrophys. J.* **208**, L47–L50 (1976).
- Meiring, J. D. et al. Elemental abundance measurements in low-redshift damped Ly $\alpha$  absorbers. *Mon. Not. R. Astron. Soc.* **370**, 43–62 (2006).
- Wolfe, A. M., Prochaska, J. X. & Gawiser, E. CII\* absorption in damped Ly $\alpha$  systems. I. Star formation rates in a two-phase medium. *Astrophys. J.* **593**, 215–234 (2003).
- Boisse, P., Le Brun, V., Bergeron, J. & Deharveng, J.-M. A HST spectroscopic study of QSOs with intermediate redshift damped Ly $\alpha$  systems. *Astron. Astrophys.* **333**, 841–863 (1998).
- Kennicutt, R. C. Jr. Star formation in galaxies along the Hubble sequence. *Annu. Rev. Astron. Astrophys.* **36**, 189–231 (1998).
- Morganti, R. et al. Neutral hydrogen in nearby elliptical and lenticular galaxies: the continuing formation of early-type galaxies. *Mon. Not. R. Astron. Soc.* **371**, 157–169 (2006).
- Le Brun, V., Bergeron, J. & Deharveng, J. M. The nature of intermediate-redshift damped Ly $\alpha$  absorbers. *Astron. Astrophys.* **321**, 733–748 (1997).
- Steidel, C. C., Pettini, M., Dickinson, M. & Persson, S. E. Imaging of two damped Lyman-alpha absorbers at intermediate redshifts. *Astron. J.* **108**, 2046–2053 (1994).
- McKee, C. F. & Ostriker, J. P. A theory of the interstellar medium: three components regulated by supernova explosions in an inhomogeneous substrate. *Astrophys. J.* **218**, 148–169 (1977).
- Morganti, R., Greenhill, L. J., Peck, A. B., Jones, D. L. & Henkel, C. Disks, tori, and cocoons: emission and absorption diagnostics of AGN environments. *N. Astron. Rev.* **48**, 1195–1209 (2004).
- Lotz, J. M. et al. The evolution of galaxy mergers and morphology at  $z \sim 1.2$  in the extended Groth strip. *Astrophys. J.* **672**, 177–197 (2008).
- Bernet, M. L., Miniati, F., Lilly, S. J., Kronberg, P. P. & Dessauges-Zavadsky, M. Strong magnetic fields in normal galaxies at high redshift. *Nature* **454**, 302–304 (2008).
- Rees, M. J. Origin of cosmic magnetic fields. *Astron. Nachr.* **327**, 395–398 (2006).
- Wolfe, A. M. & Chen, H.-W. Searching for low surface brightness galaxies in the Hubble ultra deep field: implications for the star formation efficiency in neutral gas at  $z \sim 3$ . *Astrophys. J.* **652**, 981–993 (2006).
- Heiles, C. Cross-correlation spectropolarimetry in single-dish radio astronomy. *Publ. Astron. Soc. Pacif.* **113**, 1243–1246 (2001).
- Baym, G. *Lectures on Quantum Mechanics* Ch. 1 (Benjamin, 1981).
- Savage, B. D. & Sembach, K. R. Interstellar abundances from absorption-line observations with the Hubble-Space Telescope. *Annu. Rev. Astron. Astrophys.* **34**, 279–329 (1996).

**Acknowledgements** We wish to thank F. H. Shu for suggesting the merger model and H.-W. Chen for providing us with her reanalysed images of 3C 286. We also thank F. H. Shu, E. Gawiser and A. Lazarian for comments and the US National Science Foundation for financial support. The GBT is one of the facilities of the National Radio Astronomy Observatory, which is a center of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. A.M.W., R.A.J. and J.X.P. are Visiting Astronomers at the W. M. Keck Telescope. The Keck Observatory is a joint facility of the University of California, the California Institute of Technology and the National Aeronautics and Space Administration.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.M.W. ([awolfe@ucsd.edu](mailto:awolfe@ucsd.edu)).

# Clustered star formation as a natural explanation for the H $\alpha$ cut-off in disk galaxies

Jan Pflamm-Altenburg<sup>1</sup> & Pavel Kroupa<sup>1</sup>

The rate of star formation in a galaxy is often determined by the observation of emission in the H $\alpha$  line, which is related to the presence of short-lived massive stars. Disk galaxies show a strong cut-off in H $\alpha$  radiation at a certain galactocentric distance, which has led to the conclusion that star formation is suppressed in the outer regions of disk galaxies. This is seemingly in contradiction to recent observations<sup>1</sup> in the ultraviolet which imply that disk galaxies have star formation beyond the H $\alpha$  cut-off, and that the star-formation-rate surface density is linearly related to the underlying gas surface density, which is a shallower relationship than that derived from H $\alpha$  luminosities<sup>2</sup>. In a galaxy-wide formulation, the clustered nature of star formation has recently led to the insight that the total galactic H $\alpha$  luminosity is nonlinearly related to the galaxy-wide star formation rate<sup>3</sup>. Here we show that a local formulation of the concept of clustered star formation naturally leads to a steeper radial decrease in the H $\alpha$  surface luminosity than in the star-formation-rate surface density, in quantitative agreement with the observations, and that the observed H $\alpha$  cut-off arises naturally.

The integrated galactic initial mass function (IGIMF) describes the mass spectrum of all newly formed stars in a galaxy. The IGIMF is calculated by adding all stars of all newly formed star clusters<sup>4,5</sup>, and falls off more steeply with increasing stellar masses for massive stars<sup>5</sup> than the canonical initial mass function (IMF) in each star cluster, owing to the combination of two effects: the masses of the young star clusters are distributed according to the embedded cluster mass function (ECMF), for which the upper mass limit is a function of the total star formation rate<sup>6</sup> (SFR), and the stellar upper mass limit of the IMF is a function of the total star cluster mass<sup>7</sup>. Consequently, the total fraction of massive stars and, therefore, the total H $\alpha$  luminosity decreases faster than linearly with decreasing SFR (ref. 3). The IGIMF theory has already been shown to lead naturally to the observed mass–metallicity relation of galaxies<sup>8</sup> and has received recent empirical verification in a study of IMF variations among galaxies<sup>9</sup>.

To construct a quantitative local IGIMF theory, we introduce the local embedded cluster mass function (LECMF)

$$\xi_{\text{LECMF}}(M_{\text{ecl}}, x, y) = \frac{dN_{\text{ecl}}}{dM_{\text{ecl}} dx dy}$$

which defines the number of newly formed star clusters,  $dN_{\text{ecl}}$ , in the mass interval  $[M_{\text{ecl}}, M_{\text{ecl}} + dM_{\text{ecl}}]$  per unit area at the location  $(x, y)$  in a disk galaxy. Observations<sup>10</sup> of Galactic star-forming regions show that this function is a single-part power law,  $\xi_{\text{LECMF}} \propto M_{\text{ecl}}^{-\beta}$ , with an index of  $\beta = 2$ . The least massive cluster, with mass<sup>5</sup>  $M_{\text{ecl,min}} = 5M_{\odot}$ , should form at any place in the galaxy, whereas the mass,  $M_{\text{ecl,max,loc}}(x, y)$ , of the most massive star cluster that can form locally is expected to depend on the local gas density, that is, on how much material is locally available for star cluster formation. Observations<sup>5,6</sup>

show that the mass,  $M_{\text{ecl,max}}$ , of the most massive star cluster in the whole galaxy is a function of the total galactic star formation rate. To express the dependence of the upper limit of the LECMF on the local gas surface density, we write

$$M_{\text{ecl,max,loc}}(x, y) = M_{\text{ecl,max}} \left( \frac{\Sigma_{\text{gas}}(x, y)}{\Sigma_{\text{gas,0}}} \right)^{\gamma} \quad (1)$$

where  $\Sigma_{\text{gas}}(x, y)$  and  $\Sigma_{\text{gas,0}}$  are the gas densities at the location  $(x, y)$  and at the origin, respectively.

The local mass of all star clusters between the two mass limits is determined by the local star formation rate,  $\Sigma_{\text{SFR}}(x, y)$ , which is described by the Kennicutt–Schmidt law<sup>2,11</sup>:  $\Sigma_{\text{SFR}}(x, y) = A \Sigma_{\text{gas}}^N(x, y)$ . Here  $A$  is a proportionality constant and  $N$  is widely accepted to have the value<sup>2</sup> 1.4, although  $N = 0.99$  follows from recent ultraviolet observations<sup>1</sup>. As ultraviolet emission is a star formation tracer that is much less sensitive to the presence of OB stars than H $\alpha$  emission, the true exponent  $N$  must be much closer to the value derived from ultraviolet observations. Thus, we chose  $N = 1$ . We note that the high-gas-density part of the Kennicutt–Schmidt plot<sup>2</sup> based on far-infrared observations also has a flatter slope, of  $N = 1.08$ , and that galaxy evolution models suggest  $N$  must not exceed unity if observed radial density profiles of disk galaxies are to be reproduced<sup>12</sup>, confirming our choice.

We calculate the mass spectrum of all newly formed stars per unit area—the local integrated galactic initial mass function (LIGIMF)—by adding all newly formed stars of all young star clusters, and the H $\alpha$  surface density follows by adding the H $\alpha$  flux contributions of all newly formed stars. The newly formed stars in each young star cluster are distributed according to the invariant canonical IMF (refs 13, 14) with a fixed lower mass limit but an upper mass limit depending on the total cluster mass<sup>7</sup>. In terms of H $\alpha$  emission, young star clusters above  $\sim 3,000M_{\odot}$  have constant light-to-mass ratios, whereas smaller clusters are increasingly underluminous<sup>3</sup>. With decreasing star-formation-rate surface density, the upper mass limit of the LECMF decreases and, consequently, the fraction of underluminous star clusters increases. Thus, ultraviolet and H $\alpha$  emissions scale differently with the star-formation-rate surface density, gas surface density or galactocentric radius. A detailed explanation of how the H $\alpha$  surface luminosity is calculated is given in the Supplementary Information.

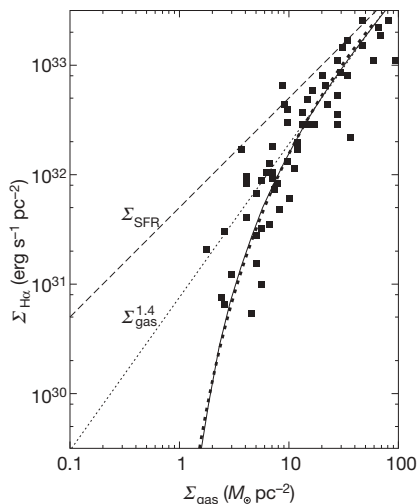
Next, we apply the LIGIMF theory to a sample of disk galaxies<sup>15</sup> with measured gas surface densities and H $\alpha$  surface luminosities of H II regions averaged over annuli at different galactocentric radii. It is known that ionising photons emitted by massive stars can escape from well-defined H II regions and lead to recombinations and, thus, H $\alpha$  radiation in the surrounding diffuse ionised gas<sup>16</sup>. Using H $\alpha$  emission as a star formation tracer, this kind of photon leakage has to be taken into account to get an estimate of the true star formation rate. A previous study<sup>16</sup> of the galaxies NGC 247 and NGC 7793 allows us to construct a correction procedure to obtain the total

<sup>1</sup>Argelander-Institut für Astronomie, Universität Bonn, 53121 Bonn, Germany.

H $\alpha$  surface luminosities from the surface luminosities of H II regions only (see Supplementary Information).

For a linear star formation law ( $N = 1$ ) as derived from ultraviolet observations<sup>1</sup>, the LIGIMF theory predicts a H $\alpha$  surface luminosity as a function of the gas surface density that is in full agreement with the observations (Fig. 1). Additionally, the radial H $\alpha$  profile derived in the LIGIMF theory matches the observations perfectly (Fig. 2). The concept of clustered star formation resolves the discrepancies between H $\alpha$  and ultraviolet observations completely.

At first sight it might be objected that the LIGIMF theory contradicts observations of the ultraviolet sources in the outer disks of galaxies: 5–10% of all clusters in the outer disks of galaxies detected in the ultraviolet have associated H $\alpha$  emission<sup>17</sup>. The age estimates of the ultraviolet knots range up to 400 Myr. Clusters with H $\alpha$  emission have ages  $\leq 20$  Myr, as they are powered by short-lived massive stars; therefore, 5% of all observed ultraviolet knots are expected to have associated H $\alpha$  emission, in agreement with observations. The LIGIMF theory predicts an overabundance of H $\alpha$ -underluminous star clusters beyond the H $\alpha$  cut-off and a smaller number ratio of H $\alpha$ -emitting to non-H $\alpha$ -emitting ultraviolet knots is expected. Underluminosity does not mean that there is no H $\alpha$  emission, however. In the LIGIMF theory, each young ultraviolet cluster is a H $\alpha$  source, too, but ultraviolet and H $\alpha$  luminosities scale differently with

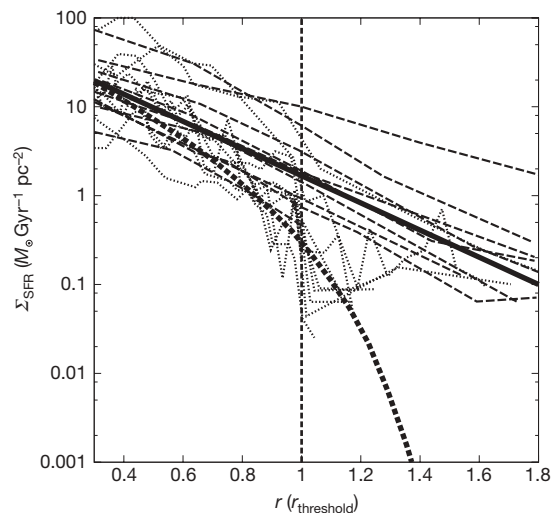


**Figure 1 | H $\alpha$  luminosity surface density ( $\Sigma_{\text{H}\alpha}$ ) versus total gas surface density ( $\Sigma_{\text{gas}}$ ).** We plot data (black squares) observed for seven disk galaxies<sup>15</sup>, averaged over annuli at different galactocentric radii, after correcting for photon leakage from H II regions (see Supplementary Information). These galaxies have a mean star formation rate of  $6.9 M_{\odot} \text{ yr}^{-1}$  ( $3.2 M_{\odot}$ – $16.4 M_{\odot} \text{ yr}^{-1}$ )<sup>2,15</sup>, a mean total gas mass of  $2.1 \times 10^{10} M_{\odot}$  ( $0.6 \times 10^{10} M_{\odot}$ – $3.6 \times 10^{10} M_{\odot}$ )<sup>2,15</sup> and a mean scale length of 4.4 kpc ( $3.9$ – $5.2$  kpc)<sup>25–28</sup>. These mean values define our model standard disk galaxy. For a choice of  $\gamma = 3/2$  (see equation (1)) the LIGIMF theory predicts a  $\Sigma_{\text{H}\alpha}$ – $\Sigma_{\text{gas}}$  relation that matches the observations excellently (solid line). We note that the underlying true star-formation-rate surface density as derived from ultraviolet observations<sup>1</sup> is directly proportional to the gas surface density ( $N = 1$ ); it is displayed after being converted into H $\alpha$  surface luminosity using the wrong linear Kennicutt H $\alpha$ –SFR relation<sup>2,29</sup> (dashed line) and shows the expected  $\Sigma_{\text{H}\alpha}$ – $\Sigma_{\text{gas}}$  relation based on the classical picture, which is in disagreement with the observations. Furthermore, the H $\alpha$  luminosity surface density in the high-luminosity part ( $\Sigma_{\text{H}\alpha} \geq 10^{32.5} \text{ erg s}^{-1} \text{ pc}^{-2}$ ) depends, for the correct LIGIMF theory, on the gas surface density raised to the power of 1.4 (dotted line, extrapolated to low H $\alpha$  surface luminosity), in agreement with the classical Kennicutt–Schmidt slope of  $N = 1.4$ . The LIGIMF theory puts the hitherto inconsistent H $\alpha$  and ultraviolet observations in perfect agreement with each other. The steeper high-luminosity slope of  $N = 1.4$  and the H $\alpha$  cut-off at low gas density are simultaneous outcomes of the LIGIMF theory. The thick dotted line which almost coincides with the thick solid curve shows a fitting function for the LIGIMF model (see Supplementary Information).

the cluster mass. Thus, this finding<sup>17</sup> is entirely consistent with the LIGIMF theory. The observed ultraviolet knots in the outer disk of the galaxy M83 are always systematically smaller than their counterparts in the inner disk<sup>18</sup>, in agreement with the fundamental basics of the LIGIMF theory. There is one outstanding massive young star cluster in the outer region of M83, but this does not contradict the theory and is instead expected from a statistical point of view (see Supplementary Information). Furthermore, the M83 far-ultraviolet luminosity function of outer-disk stellar complexes is steeper than that of the inner-disk population<sup>18</sup>.

A similar trend is reported in the galaxy NGC 628 for the H $\alpha$  luminosity function of H II regions<sup>19</sup>. In the LIGIMF theory, inner-disk LECMFs have higher upper mass limits than outer-disk LECMFs. Integration of the LECMFs over the outer and inner regions leads to an outer-disk ECMF steeper than the inner-disk ECMF, indicating that outer-disk star formation complexes are systematically less massive than those in the inner disk. This integration effect is fundamentally equivalent to the IGIMF being steeper for dwarf galaxies with low global star formation rates than for disk galaxies with high star formation rates<sup>3</sup>.

Previously, the H $\alpha$  cut-off has been explained<sup>15</sup> by a drop of the local gas density below a critical value determined by the stability condition of a thin isothermal disk<sup>20,21</sup> where no star formation can occur. In contradiction to this explanation, recent ultraviolet observations<sup>1</sup> reveal star formation outside the H $\alpha$  cut-off, and dwarf galaxies<sup>22</sup> show star formation although their average gas density is lower than the critical value. Indeed, it has been shown that in regions with densities lower than the critical value, star formation can be driven by instabilities other than thermal<sup>23</sup>. It has been argued that H II regions powered by the same massive stars are larger in a thin environment—that is, at large galactocentric radii—than in a dense one, and identical H II regions thus become fainter in the outer galaxy. Therefore, it has been concluded<sup>23</sup> that the H $\alpha$  surface luminosity should decrease faster than the star-formation-rate surface density. Indeed, the surface brightness of individual H II regions should



**Figure 2 | Star-formation-rate surface density ( $\Sigma_{\text{SFR}}$ ) versus galactocentric radius ( $r$ ).** Radial distribution of the star-formation-rate surface density of nine disk galaxies based on ultraviolet<sup>1</sup> (thin dash-dot lines) and H $\alpha$  (ref. 30; thin dotted lines) observations that rely on a wrong linear conversion<sup>29</sup> between the corresponding H $\alpha$  luminosity surface density and star-formation-rate surface density after correction for photon leakage (see Supplementary Information). The galactocentric radius,  $r$ , is expressed in units of  $r_{\text{threshold}}$ , the H $\alpha$  threshold radius<sup>30</sup>. Also plotted are the true underlying star-formation-rate surface density of our standard disk galaxy (thick solid line) as defined in Fig. 1 and the model H $\alpha$  surface luminosity (thick dotted line) converted into a star-formation-rate surface density using the same linear conversion<sup>29,30</sup>. The LIGIMF theory thus naturally accounts for the discrepant  $\Sigma_{\text{SFR}}$  value at a particular radius.



be fainter in the outer galaxy. However, the  $H\alpha$  surface density considered in star formation laws refers to the total  $H\alpha$  luminosity per unit area of the galaxy and not to the cross-section of the  $H\text{ II}$  region. Identically powered  $H\text{ II}$  regions contribute equally to the  $H\alpha$  surface luminosity independently of their location in a thin or a dense gas environment. Thus, the proposed solution<sup>23</sup> explains neither the  $H\alpha$  cut-off nor the different slopes of the ultraviolet-based and  $H\alpha$ -based star formation laws. It has been shown recently that a required minimum column density for massive star formation might exist<sup>24</sup>, implying star formation with no massive stars in low-density environments. However, this model predicts a top-heavy IMF for cloud column densities much larger than this threshold, for which no observational evidence exists<sup>14</sup>, and allows no quantitative linkage of  $H\alpha$  luminosity and the star formation rate.

Contrary to this previously existing work, the LIGIMF theory developed here is in excellent agreement with the observed radial  $H\alpha$  and ultraviolet luminosity profiles (Fig. 2) and the Kennicutt–Schmidt star formation law (Fig. 1), and also allows the determination of star formation rates even in  $H\alpha$ -faint galaxy regions.

Received 6 March; accepted 14 July 2008.

- Boissier, S. et al. Radial variation of attenuation and star formation in the largest late-type disks observed with GALEX. *Astrophys. J. Suppl. Ser.* **173**, 524–537 (2007).
- Kennicutt, R. C. Jr. The global Schmidt law in star-forming galaxies. *Astrophys. J.* **498**, 541–552 (1998).
- Pflamm-Altenburg, J., Weidner, C. & Kroupa, P. Converting  $H\alpha$  luminosities into star formation rates. *Astrophys. J.* **671**, 1550–1558 (2007).
- Vanbeveren, D. Theoretical evolution of massive stellar aggregates. *Astron. Astrophys.* **124**, 71–76 (1983).
- Weidner, C. & Kroupa, P. The variation of integrated star initial mass functions among galaxies. *Astrophys. J.* **625**, 754–762 (2005).
- Weidner, C., Kroupa, P. & Larsen, S. S. Implications for the formation of star clusters from extragalactic star formation rates. *Mon. Not. R. Astron. Soc.* **350**, 1503–1510 (2004).
- Weidner, C. & Kroupa, P. Evidence for a fundamental stellar upper mass limit from clustered star formation. *Mon. Not. R. Astron. Soc.* **348**, 187–191 (2004).
- Köppen, J., Weidner, C. & Kroupa, P. A possible origin of the mass-metallicity relation of galaxies. *Mon. Not. R. Astron. Soc.* **375**, 673–684 (2007).
- Hoversten, E. A. & Glazebrook, K. Evidence for a nonuniversal stellar initial mass function from the integrated properties of SDSS galaxies. *Astrophys. J.* **675**, 163–187 (2008).
- Lada, C. J. & Lada, E. A. Embedded clusters in molecular clouds. *Annu. Rev. Astron. Astrophys.* **41**, 57–115 (2003).
- Kennicutt, R. C. Jr et al. Star formation in NGC 5194 (M51a). II. The spatially resolved star formation law. *Astrophys. J.* **671**, 333–348 (2007).
- Zasov, A. V. & Abramova, O. V. The star-formation efficiency and density of the disks of spiral galaxies. *Astron. Rep.* **50**, 874–886 (2006).
- Kroupa, P. On the variation of the initial mass function. *Mon. Not. R. Astron. Soc.* **322**, 231–246 (2001).
- Kroupa, P. The initial mass function of stars: Evidence for uniformity in variable systems. *Science* **295**, 82–91 (2002).
- Kennicutt, R. C. Jr. The star formation law in galactic disks. *Astrophys. J.* **344**, 685–703 (1989).
- Ferguson, A. M. N., Wyse, R. F. G., Gallagher, J. S. III & Hunter, D. A. Diffuse ionized gas in spiral galaxies: Probing Lyman continuum photon leakage from  $H\text{ II}$  regions? *Astron. J.* **111**, 2265–2279 (1996).
- Zaritsky, D. & Christlein, D. On the extended knotted disks of galaxies. *Astron. J.* **134**, 135–141 (2007).
- Thilker, D. A. et al. Recent star formation in the extreme outer disk of M83. *Astrophys. J.* **619**, L79–L82 (2005).
- Lelièvre, M. & Roy, J.-R. The  $H\text{ II}$  regions of the extreme outer disk of NGC 628. *Astron. J.* **120**, 1306–1315 (2000).
- Toomre, D. A. On the gravitational stability of a disk of stars. *Astrophys. J.* **139**, 1217–1238 (1964).
- Cowie, L. L. Cloud fluid compression and softening in spiral arms and the formation of giant molecular cloud complexes. *Astrophys. J.* **245**, 66–71 (1981).
- Hunter, D. A., Elmegreen, B. G. & van Woerden, H. Neutral hydrogen and star formation in the irregular galaxy NGC 2366. *Astrophys. J.* **556**, 773–800 (2001).
- Elmegreen, B. G. & Hunter, D. A. Radial profiles of star formation in the far outer regions of galaxy disks. *Astrophys. J.* **636**, 712–720 (2006).
- Krumholz, M. R. & McKee, C. F. A minimum column density of  $1\text{ g cm}^{-2}$  for massive star formation. *Nature* **451**, 1082–1084 (2008).
- Kenney, J. D. P. & Young, J. S. The effects of environment on the molecular and atomic gas properties of large Virgo cluster spirals. *Astrophys. J.* **344**, 171–199 (1989).
- Wong, T. & Blitz, L. The relationship between gas content and star formation in molecule-rich spiral galaxies. *Astrophys. J.* **569**, 157–183 (2002).
- Schuster, K. F., Kramer, C., Hitschfeld, M., Garcia-Burillo, S. & Mookerjee, B. A complete  $^{12}\text{CO}$  2–1 map of M 51 with HERA. I. Radial averages of CO, HI, and radio continuum. *Astron. Astrophys.* **461**, 143–151 (2007).
- Crosthwaite, L. P. & Turner, J. L. CO(1–0), CO(2–1), and neutral gas in NGC 6946: Molecular gas in a late-type, gas-rich, spiral galaxy. *Astron. J.* **134**, 1827–1842 (2007).
- Kennicutt, R. C. Jr, Tamblyn, P. & Congdon, C. E. Past and future star formation in disk galaxies. *Astrophys. J.* **435**, 22–36 (1994).
- Martin, C. L. & Kennicutt, R. C. Jr. Star formation thresholds in galactic disks. *Astrophys. J.* **555**, 301–321 (2001).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank K. S. de Boer for discussions.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.P.-A. ([jpflamm@astro.uni-bonn.de](mailto:jpflamm@astro.uni-bonn.de)).

## LETTERS

# Nanoscale magnetic sensing with an individual electronic spin in diamond

J. R. Maze<sup>1</sup>, P. L. Stanwix<sup>2</sup>, J. S. Hodges<sup>1,3</sup>, S. Hong<sup>1</sup>, J. M. Taylor<sup>4</sup>, P. Cappellaro<sup>1,2</sup>, L. Jiang<sup>1</sup>, M. V. Gurudev Dutt<sup>5</sup>, E. Togan<sup>1</sup>, A. S. Zibrov<sup>1</sup>, A. Yacoby<sup>1</sup>, R. L. Walsworth<sup>1,2</sup> & M. D. Lukin<sup>1</sup>

Detection of weak magnetic fields with nanoscale spatial resolution is an outstanding problem in the biological and physical sciences<sup>1–5</sup>. For example, at a distance of 10 nm, the spin of a single electron produces a magnetic field of about 1  $\mu$ T, and the corresponding field from a single proton is a few nanoteslas. A sensor able to detect such magnetic fields with nanometre spatial resolution would enable powerful applications, ranging from the detection of magnetic resonance signals from individual electron or nuclear spins in complex biological molecules<sup>5,6</sup> to readout of classical or quantum bits of information encoded in an electron or nuclear spin memory<sup>7</sup>. Here we experimentally demonstrate an approach to such nanoscale magnetic sensing, using coherent manipulation of an individual electronic spin qubit associated with a nitrogen-vacancy impurity in diamond at room temperature<sup>8</sup>. Using an ultra-pure diamond sample, we achieve detection of 3 nT magnetic fields at kilohertz frequencies after 100 s of averaging. In addition, we demonstrate a sensitivity of 0.5  $\mu$ T Hz<sup>–1/2</sup> for a diamond nanocrystal with a diameter of 30 nm.

Sensitive solid-state magnetometers typically use phenomena such as superconducting quantum interference in SQUIDS<sup>2,3</sup> or the Hall effect in semiconductors<sup>4</sup>. Intriguing avenues such as magnetic resonance force microscopy are also currently being explored<sup>5,6</sup>. Our approach to magnetic sensing<sup>8</sup> uses the coherent manipulation of a single quantum system, an electronic spin qubit. As illustrated in Fig. 1, the electronic spin of an individual nitrogen-vacancy impurity in diamond can be polarized by optical pumping and measured through state-selective fluorescence. Conventional electron spin resonance (ESR) techniques are used to coherently manipulate its orientation. To achieve magnetic sensing, we monitor the electronic spin precession, which depends on external magnetic fields through the Zeeman effect. This method is directly analogous to precision measurement techniques in atomic and molecular systems<sup>9</sup>, which are widely used to implement ultra-stable atomic clocks<sup>10–12</sup> and sensitive magnetometers<sup>13</sup>.

The principal challenge for achieving high sensitivity using solid-state spins is their strong coupling to the local environment, which limits the free precession time and thus the magnetometer's sensitivity. Recently, there has been great progress in understanding the local environment of nitrogen-vacancy spin qubits, including <sup>13</sup>C nuclear spins<sup>7,14–17</sup> and electronic spin impurities<sup>18–20</sup>. Here we use coherent control over a coupled electron–nuclear system<sup>8,16</sup>, similar to techniques used in magnetic resonance, to decouple the magnetometer spin from its environment. As illustrated in Fig. 1d, a spin-echo sequence refocuses the unwanted evolution of the magnetometer spin due to environmental fields fluctuating randomly on timescales much longer than the length of the sequence. However, oscillating

external magnetic fields matching the echo period will affect the spin dynamics constructively, allowing sensitive detection of its amplitude.

The ideal preparation, manipulation and detection of an electronic spin would yield a so-called quantum-projection-noise-limited minimum detectable magnetic field<sup>12</sup>

$$\delta B_{\min} \approx \frac{\hbar}{g\mu_B\sqrt{T_2T}} \quad (1)$$

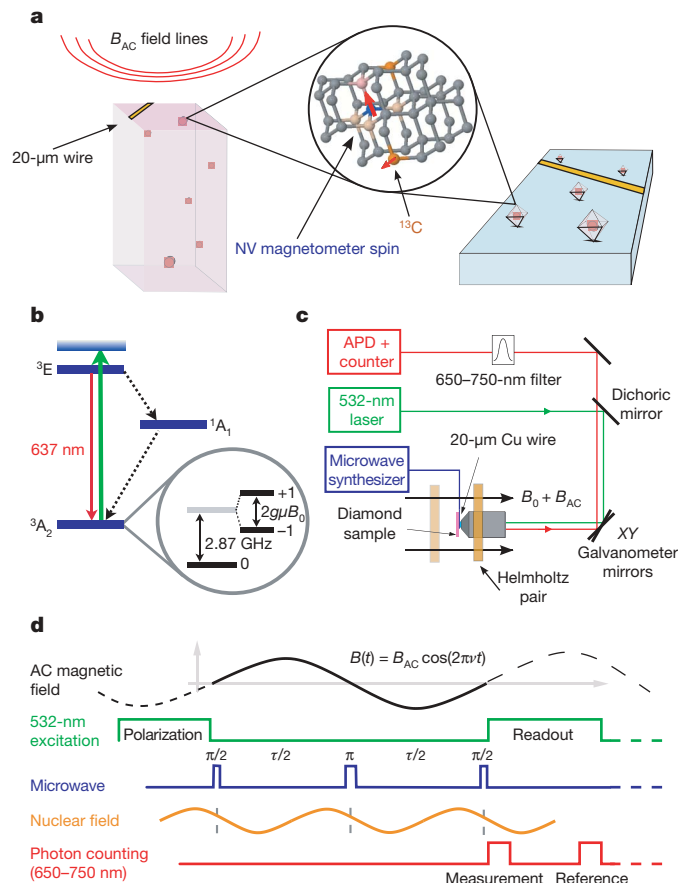
where  $T_2$  is the electronic spin coherence time,  $T$  is the measurement time,  $\mu_B$  is the Bohr magneton,  $\hbar$  is Planck's constant divided by  $2\pi$ , and  $g \approx 2$  is the electronic Landé  $g$ -factor. In principle, for typical values of  $T_2 \approx 0.1$ –1 ms, sensitivity of the order of a few nT Hz<sup>–1/2</sup> can be achieved with a single nitrogen-vacancy centre. Although this is less sensitive than for state-of-the-art macroscopic magnetometers<sup>1,3</sup>, a key feature of our sensor is that it can be localized within a region of about 10 nm, either in direct proximity to a diamond surface or within a nano-sized diamond crystal (Fig. 1a). Sensitive magnetic detection on a nanometre scale can then be performed with such a system under ambient conditions. Supplementary Fig. 1 provides a comparison between magnetic field sensitivity and detector volume for several state-of-the-art magnetometers and the nitrogen-vacancy diamond systems demonstrated here.

To establish the sensitivity limits of a single electronic spin magnetometer, we carried out a series of proof-of-principle experiments involving single nitrogen-vacancy centres in bulk ultra-pure single-crystal diamond and in commercially available diamond nanocrystals. Our experimental methodology is outlined schematically in Fig. 1; further details about our experimental set-up and diamond samples are given in Methods. We first focus on the single-crystal diamond bulk sample. Figure 2a shows a typical spin-echo signal observed from an individual nitrogen-vacancy centre. The periodic modulation of the echo is caused by a bath of spin-1/2 <sup>13</sup>C nuclei (1.1% natural abundance), which create an effective precessing magnetic field at the nitrogen-vacancy centre of a few microteslas. In the presence of an applied static magnetic field  $B_{DC}$ , the periodic Larmor precession of the nuclear field causes the nitrogen-vacancy spin-echo signal to collapse and revive<sup>16</sup> at half the rate of the Larmor frequency of <sup>13</sup>C,  $\omega_L = \gamma_{13C} B_{DC}$ , where  $\gamma_{13C}$  is the carbon gyromagnetic ratio. Note that substantial spin-echo revivals exist even after a free evolution of 0.6 ms. To detect an external AC magnetic field with the highest sensitivity, we must eliminate the contribution from the <sup>13</sup>C nuclear field. To this end, the revival rate of the spin-echo signal is adjusted by varying the strength of  $B_{DC}$ , such that the frequency of the echo revival peaks coincides with multiples of the AC field frequency ( $\nu$ ) to be detected.

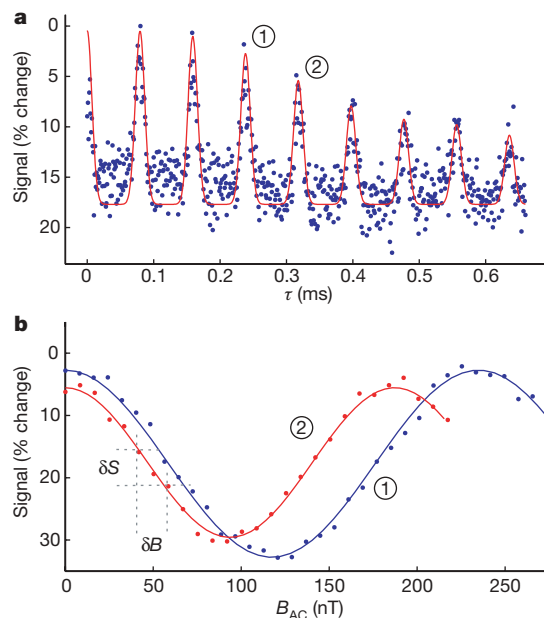
<sup>1</sup>Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>2</sup>Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA.

<sup>3</sup>Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>4</sup>Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02138, USA. <sup>5</sup>Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.

As shown in Fig. 2b, the observed peak of the spin-echo signal varies periodically as the amplitude of the external AC field ( $B_{AC}$ ) is increased. This signal variation results from phase accumulated by the nitrogen-vacancy spin due to the external AC magnetic field and the resultant time-varying Zeeman shift during the spin's precession; converting this phase into a spin population difference gives rise to variations in the detected fluorescence, which serves as the magnetometer signal. Note that the period of this signal oscillation depends on the spin-echo interval,  $\tau = 1/\nu$ . For a given value of  $B_{AC}$ , the phase accumulated by the electronic spin over one period will increase as



**Figure 1 | Principles of the magnetic sensor, which is based on individual nitrogen-vacancy electronic spins in diamond.** **a**, A single nitrogen-vacancy impurity (NV) proximal to the surface of an ultra-pure bulk single-crystal diamond sample (left) or localized within a diamond nanocrystal (right) is used to sense an externally applied AC magnetic field ( $B_{AC}$ , top left). A 20- $\mu\text{m}$ -diameter wire (yellow) generates microwave pulses to manipulate the electronic spin states. **b**, Level structure of the nitrogen-vacancy centre; see Methods for details. **c**, Diagram of the experimental approach. Single nitrogen-vacancy centres are imaged and localized with  $\sim 170$  nm resolution using confocal microscopy. The position of the focal point is moved near the sample surface, using a galvanometer mounted mirror to change the beam path and a piezo-driven objective mount. A pair of Helmholtz coils is used to provide both AC and DC magnetic fields. Experiments are then performed on single nitrogen-vacancy centres, as verified by photon correlation measurements. **d**, Optical and microwave spin-echo pulse sequence used for sensing an AC magnetic field,  $B_{AC}(\tau)$ . An individual centre is first polarized into the  $m_s = 0$  sublevel. A coherent superposition between the states  $m_s = 0$  and  $m_s = 1$  is created by applying a microwave  $\pi/2$  pulse tuned to this transition. The system freely evolves for a period of time  $\tau/2$ , followed by a  $\pi$  refocusing pulse. After a second  $\tau/2$  evolution period, the electronic spin state is projected onto the  $m_s = 0, 1$  basis by a final  $\pi/2$  pulse, at which point the ground state population is detected optically via spin-dependent fluorescence. The DC magnetic field is adjusted to eliminate the contribution of the randomly phased field produced by  $^{13}\text{C}$  nuclear spins (gold curve) by choosing  $\tau = 2n/\omega_L$ , for integer  $n$ .



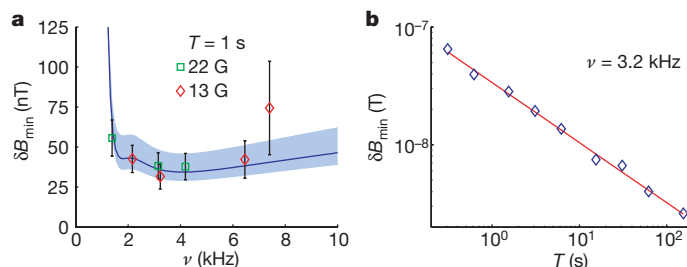
**Figure 2 | Demonstration of spin-echo-based magnetometry with an individual nitrogen-vacancy electronic spin in a bulk diamond sample.** **a**, Example of electronic spin-echo measurement. We plot the normalized echo signal corresponding to a fractional change of nitrogen-vacancy centre fluorescence. Maximal signal corresponds to an average number of photons  $\langle n \rangle = 0.03$  detected during the 324-ns photon counting window of a single experimental run. Collapses and revivals are due to interactions with a  $^{13}\text{C}$  nuclear spin bath. The revivals occur at half the rate of the Larmor frequency of  $^{13}\text{C}$  (here set by  $B_{DC} = 22$  G). The spin-echo signal envelope was fitted with an exponential decay function modulated by a strongly interacting pair of nearby  $^{13}\text{C}$  (see Methods). Magnetometer sensitivity experiments are performed at spin-echo revival peaks to maximize signal. Revivals 1 and 2, treated in **b**, are indicated. **b**, Examples of measured spin-echo signal as a function of  $B_{AC}$  for two operating frequencies,  $\nu_1 = 3.15$  kHz (red) and  $\nu_2 = 4.21$  kHz (blue), corresponding to revivals 1 and 2 indicated in **a**. Each displayed point is a result of  $N = 7 \times 10^5$  averages of spin-echo sequences. The magnetometer is most sensitive to variations in the AC magnetic field amplitude ( $\delta B$ ) at the point of maximum slope, with the sensitivity being limited by the uncertainty in the spin-echo signal measurement ( $\delta S$ ). We note that the cosine behaviour of the signal with respect to AC magnetic field amplitude can be changed to a sine by adjusting the phase of the third microwave pulse by  $90^\circ$ . This change moves the point of maximum magnetometer sensitivity to near zero AC field amplitude.

the frequency of the external AC field decreases. At the conclusion of a single run of the magnetometry pulse sequence, the measurable spin-echo signal  $S_B$  is proportional to the probability of the nitrogen-vacancy spin being in the  $m_s = 0$  state:  $S_B \propto P_0(B_{AC}) = [1 + F(\tau)\cos(\delta\phi)]/2$ , where  $\delta\phi = 4g\mu_B B_{AC}/2\pi\nu$  and  $F(\tau)$  is the amplitude of the spin-echo signal envelope in the absence of the external AC magnetic field (Fig. 2a).

The sensitivity of the nitrogen-vacancy magnetometer to small variations in  $B_{AC}$ , as depicted in the measurements shown in Fig. 2b, is given by  $\delta B_{\min} = \sigma_S^N / dS_B$ , where  $\sigma_S^N$  is the standard deviation of the spin-echo measurement after  $N$  averages and  $dS_B$  is the slope of the spin-echo signal variation with  $B_{AC}$ . Since maximum sensitivity (that is, smallest  $\delta B_{\min}$ ) occurs at maximum slope, all magnetometer sensitivity measurements were conducted at this point. This maximum slope is proportional to the spin-echo amplitude divided by the frequency of the oscillating field,  $dS_B \propto F(1/\nu)/\nu$ . For a shot-noise-limited signal with uncertainty  $\sigma_S$  in a single measurement,  $\sigma_S^N = \sigma_S/\sqrt{N}$ , where  $N = T/\tau$ . Hence the magnetometer sensitivity is expected to scale as  $\delta B_{\min} \propto \sqrt{\nu}/F(1/\nu)$ .

Figure 3a shows example measurements of the sensitivity  $\delta B_{\min}$  after 1 s of averaging as a function of the AC magnetic field frequency,  $\nu = 1/\tau$ . As this frequency decreases, the accumulated Zeeman phase





**Figure 3 | Characterization of magnetometer sensitivity and minimum measurable AC magnetic field.** **a**, Measured sensitivity of a single nitrogen-vacancy spin magnetometer in a bulk diamond sample over a range of frequencies for the external AC magnetic field after averaging for one second ( $T = 1$  s). Error bars, standard deviation (s.d.) for a sample size of 30. Also shown is the theoretically predicted sensitivity (solid blue line), with the shaded region representing uncertainty due to variations in photon

shift of the nitrogen-vacancy spin during one period increases. This makes the nitrogen-vacancy spin more sensitive to variations of  $B_{AC}$  as the frequency is reduced, until the point at which the nitrogen-vacancy spin decoheres during a single period of the external AC magnetic field's oscillation. This decoherence decreases the magnetometer's sensitivity by decreasing the contrast of the spin-echo signal ( $F(1/\nu) \rightarrow 0$ ) and therefore the slope  $dS_B$ . At high frequencies or short times,  $F(1/\nu) \rightarrow 1$ , and the sensitivity scales as  $\sqrt{\nu}$ . Hence, the magnetometer sensitivity is optimized for frequencies comparable with the longest time for which substantial echo signal is still observable. We note that it is possible to measure at higher frequencies without further loss of sensitivity by using multiple spin-echo pulses in a given measurement period<sup>8</sup>. Figure 3b shows examples of measured nitrogen-vacancy magnetometer sensitivity for a fixed  $\nu$  as a function of  $T$ . The solid line is a fit to  $\delta B_{\min} \propto T^{-\alpha}$ , where  $\alpha = 0.5 \pm 0.01$ , indicating that magnetic fields as small as few nanoteslas are resolvable after 100 s of averaging.

As noted above, a key feature of our technique is that at specific times, determined by echo revivals, the nitrogen-vacancy electronic spin can be essentially decoupled from  $^{13}\text{C}$  nuclear spins. In practice, the decoupling is not perfect, owing to internal dynamics of the electronic environment other than simple spin precession. In fact, the overall decay of the echo signal shown in Fig. 2a does not follow the simple exponential decay associated with typical ESR on bulk samples. This can be understood by noting that the echo dynamics of a single nitrogen-vacancy centre near its revivals is probably determined by a few nearby  $^{13}\text{C}$  atoms, which interact strongly with the electronic spin<sup>7,14–16,21</sup>, yielding multiple characteristic timescales for echo decay (see Methods).

The absolute sensitivity of the nitrogen-vacancy magnetometer depends on the signal-to-noise ratio in the readout of the nitrogen-vacancy electronic spin state. In the present demonstration, this is limited by photon collection efficiency, which is  $\sim 0.1\%$ . The resulting photon shot noise<sup>1,8</sup> is about an order of magnitude larger than the ideal quantum projection noise limit given by equation (1), resulting in a corresponding degradation of magnetometer sensitivity. Our theoretical prediction of magnetometer sensitivity (solid curve in Fig. 3a) combines the nitrogen-vacancy coherence properties shown in Fig. 2a with the noise due to photon counting statistics and imperfect collection efficiency (see Methods). This prediction is in excellent agreement with our experimental results, indicating that our magnetometer is photon-shot-noise limited.

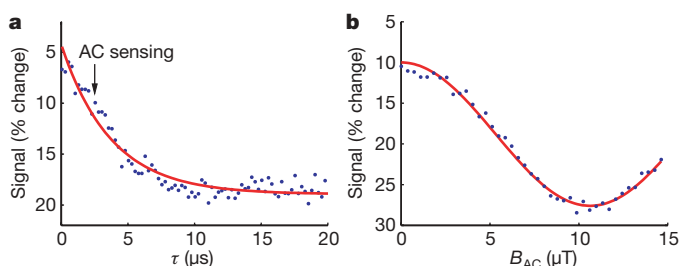
To demonstrate magnetic sensing within a nanoscale detection volume, we also performed similar experiments with single nitrogen-vacancy centres in diamond nanocrystals. We used commercially available nanocrystals that contain a large number of impurities, which shorten the electronic spin coherence time<sup>22</sup> to values ranging from 4 to 10  $\mu\text{s}$ . Sensitive detection of AC magnetic fields is still possible, as demonstrated experimentally in Fig. 4. Here,

collection efficiency (see Methods). Measurements were carried out at two different DC fields,  $B_{DC} = 13$  G (in red) and 22 G (in green). **b**, The minimum measurable AC magnetic field as a function of averaging time, for AC field frequency  $\nu = 3.2$  kHz and  $B_{DC} = 13$  G. Fit to this data (red curve) shows that the sensitivity improves as the square root of the averaging time, and is consistent with theoretical estimates based on photon-shot-noise limited detection.

the echo signal from a single nitrogen-vacancy centre in a 30-nm-size nanocrystal decays on a timescale of  $\sim 4$   $\mu\text{s}$ . The absence of characteristic collapses and revivals, associated with couplings to  $^{13}\text{C}$  nuclear spins, indicates that the echo decay is probably due to other spin impurities, such as paramagnetic substitutional nitrogen atoms containing unpaired electron spins. Magnetic sensing with such a nanocrystal at  $\nu = 380$  kHz is demonstrated in Fig. 4b. From these measurements, we estimate a magnetometer sensitivity of  $\delta B_{\min} \approx 0.5 \pm 0.1$   $\mu\text{T Hz}^{-1/2}$  for this nanocrystal.

Improved magnetometer sensitivity for bulk and nanocrystal diamond may be achieved in several ways. By using isotopically pure diamond with low concentrations of both  $^{13}\text{C}$  and nitrogen electron spin impurities, much longer coherence and interrogation times should be possible. For diamond nanocrystals, however, the ultimate sensitivity will eventually be limited by surface effects<sup>19,23</sup>. Increases to the signal-to-noise ratio may also be possible by improving the measurement readout efficiency. Near single-shot readout of an electronic spin in diamond has been achieved with cryogenic cooling using resonant excitation<sup>24</sup>. Photon collection efficiency at room temperature can also be substantially improved using either conventional far-field optics or evanescent, near-field coupling to optical waveguides<sup>25</sup>. Finally, further improvements can probably be obtained by using magnetic sensing with multiple nitrogen-vacancy centres and by using more complex pulse sequences<sup>8</sup>.

Our results demonstrate that electronic spins in diamond can be used for precision measurements of nanoscale magnetic fields. This approach opens a new regime of magnetic sensing, enabling detection of single-electron and even nuclear spins separated from



**Figure 4 | Demonstration of magnetic sensing with a single nitrogen-vacancy electronic spin in a diamond nanocrystal.** **a**, Example of electronic spin-echo signal from a single nitrogen-vacancy centre contained in a diamond nanocrystal with diameter of  $34 \pm 12$  nm as determined by atomic force microscopy. Maximum signal corresponds to an average number of photons  $\langle n \rangle = 0.02$  counted during a 324-ns photon counting window. The arrow indicates the time at which magnetic sensing is performed in **b**. **b**, Example of electronic spin-echo signal as a function of  $B_{AC}$  at a frequency  $\nu = 380$  kHz. For these data,  $N = 2 \times 10^6$  averages of spin-echo sequences were used. The resulting standard deviation yields a magnetometer sensitivity of  $0.5 \pm 0.1$   $\mu\text{T Hz}^{-1/2}$ .

nitrogen-vacancy centres by a few tens of nanometres (see Supplementary Information for details). For example, by combining our spin-echo based method with the recently demonstrated<sup>26</sup> transport and manipulation of nanocrystals using an atomic force microscope, a new kind of nanoscale scanning magnetic sensor may be created. Such a sensor could have a wide range of applications, ranging from biological and materials science to quantum information processing and fundamental tests of quantum mechanics. With the aid of field gradients, used for example in approaches based on magnetic resonance force microscopy<sup>5,6</sup>, nitrogen-vacancy diamond magnetometers may allow sensing and resolving of individual nuclear spins, with applications in structural biology<sup>8,27</sup>. Our sensing technique also provides an efficient method for measuring single electronic spins in various quantum computing architectures. Furthermore, this technique may allow non-destructive mapping of quantum states into nitrogen-vacancy centres, operating as a quantum magnetic 'head'<sup>28</sup>, with possibilities for mechanical transport of quantum information. Finally, we note that our technique could be used for detecting the quantum motion of magnetic mechanical resonators<sup>29,30</sup>, with new possibilities for creating non-classical states of mechanical motion and for testing quantum mechanics on a macroscopic scale.

## METHODS SUMMARY

AC magnetometry was performed at room temperature on nitrogen-vacancy centres found in both a bulk single-crystal diamond sample and in synthetic diamond nanocrystals (30 nm mean diameter). Single nitrogen-vacancy centres were isolated and probed by confocal microscopy. Phonon-mediated fluorescent emission (630–750 nm) was detected under coherent optical excitation ( $\lambda = 532$  nm) using a single photon counting module (APD). As single spots in the confocal image may constitute many nitrogen-vacancy centres, single centres were identified by observing photon antibunching in the measurement of the second-order correlation function.

Green excitation of a nitrogen-vacancy centre also polarized the electronic spin by optical pumping to the  $m_s = 0$  sublevel of the  $^3A_2$  ground state. The mechanism responsible for optical pumping also provided a means for spin-sensitive detection, as the rate of fluorescence differs for the  $m_s = 0$  and  $m_s = \pm 1$  states. Coherent manipulation of the spin states was achieved by applying microwave radiation resonant with the  $|0\rangle \rightarrow |1\rangle$  transition through a 20  $\mu\text{m}$  wire. A pair of Helmholtz coils provided a static magnetic field to split the degenerate  $|\pm 1\rangle$  levels; these coils also produced the external AC magnetic fields sensed with the nitrogen-vacancy magnetometer.

In performing magnetometry, pulsed laser and microwave excitations were defined with an acousto-optic modulator and microwave switch, respectively. As described in Fig. 1d, magnetometer measurements were made for an external AC magnetic field with amplitude  $B_{AC}$  and frequency  $\nu$ , properly phased with respect to the microwave pulses. When the length of the spin-echo sequence ( $\tau$ ) equalled  $1/\nu$ , the accumulated phase of the electronic spin was proportional to  $B_{AC}$ . The fluorescence rate was directly related to this phase. A counting window of 324 ns provided optimal contrast of the fluorescent readout. Many spin-echo cycles were typically averaged to reduce the uncertainty of the photon statistics associated with the low count rate ( $<1$  photon per readout). This technique was sensitive to the projection of the AC magnetic field onto the quantization axis of the electronic spin, corresponding to a vector magnetometer.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 23 April; accepted 18 July 2008.**

1. Budker, D. & Romalis, M. Optical magnetometry. *Nature Phys.* **3**, 227–234 (2007).
2. Bending, S. J. Local magnetic probes of superconductors. *Adv. Phys.* **48**, 449–535 (1999).
3. Kleiner, R., Koelle, D., Ludwig, F. & Clarke, J. Superconducting quantum interference devices: State of the art and applications. *Proc. IEEE* **92**, 1534–1548 (2004).

4. Owston, C. N. A Hall effect magnetometer for small magnetic fields. *J. Sci. Instrum.* **44**, 798–800 (1967).
5. Rugar, D., Budakian, R., Mamin, H. J. & Chui, B. W. Single spin detection by magnetic resonance force microscopy. *Nature* **430**, 329–332 (2004).
6. Mamin, H. J., Poggio, M., Degen, C. L. & Rugar, D. Nuclear magnetic resonance imaging with 90-nm resolution. *Nature Nanotechnol.* **2**, 301–306 (2007).
7. Dutt, M. V. G. *et al.* Quantum register based on individual electronic and nuclear spin qubits in diamond. *Science* **316**, 1312–1316 (2007).
8. Taylor, J. *et al.* High-sensitivity diamond magnetometer with nanoscale resolution. *Nature Phys.* (in the press); preprint at (<http://arXiv.org/abs/0805.1367v1>) (2008).
9. Budker, D. F., Kimball, D. F. & DeMille, D. P. *Atomic Physics: An Exploration Through Problems and Solutions* (Oxford Univ. Press, 2004).
10. Ludlow, A. D. *et al.* Sr lattice clock at  $1 \times 10^{-16}$  fractional uncertainty by remote optical evaluation with a Ca clock. *Science* **319**, 1805–1808 (2008).
11. Rosenband, T. *et al.* Frequency ratio of  $\text{Al}^{+}$  and  $\text{Hg}^{+}$  single-ion optical clocks; metrology at the 17th decimal place. *Science* **319**, 1808–1812 (2008).
12. Wineland, D. J., Bollinger, J. J., Itano, W. M., Moore, F. L. & Heinzen, D. J. Spin squeezing and reduced quantum noise in spectroscopy. *Phys. Rev. A* **46**, R6797 (1992).
13. Kominis, I. K., Kornack, T. W., Allred, J. C. & Romalis, M. V. A subfemtotesla multichannel atomic magnetometer. *Nature* **422**, 596–599 (2003).
14. Jelezko, F., Gaebel, T., Popa, I., Gruber, A. & Wrachtrup, J. Observation of coherent oscillations in a single electron spin. *Phys. Rev. Lett.* **92**, 076401 (2004).
15. Jelezko, F. *et al.* Observation of coherent oscillation of a single nuclear spin and realization of a two-qubit conditional quantum gate. *Phys. Rev. Lett.* **93**, 130501 (2004).
16. Childress, L. *et al.* Coherent dynamics of coupled electron and nuclear spin qubits in diamond. *Science* **314**, 281–285 (2006).
17. Jiang, L. *et al.* Coherence of an optically illuminated single nuclear spin qubit. *Phys. Rev. Lett.* **100**, 073001 (2008).
18. Hanson, R., Mendoza, F. M., Epstein, R. J. & Awschalom, D. D. Polarization and readout of coupled single spins in diamond. *Phys. Rev. Lett.* **97**, 087601 (2006).
19. Gaebel, T. *et al.* Room-temperature coherent coupling of single spins in diamond. *Nature Phys.* **2**, 408–413 (2006).
20. Hanson, R., Dobrovitski, V. V., Feiguin, A. E., Gywat, O. & Awschalom, D. D. Coherent dynamics of a single spin interacting with an adjustable spin bath. *Science* **320**, 352–355 (2008).
21. Maze, J. R., Taylor, J. M. & Lukin, M. D. Electron spin decoherence of single nitrogen-vacancy defects in diamond. Preprint at (<http://arXiv.org/abs/0805.0327>) (2008).
22. Rabeau, J. R. *et al.* Single nitrogen vacancy centers in chemical vapor deposited diamond nanocrystals. *Nano Lett.* **7**, 3433–3437 (2007).
23. Rabeau, J. R. *et al.* Implantation of labelled single nitrogen vacancy centers in diamond using  $^{15}\text{N}$ . *Appl. Phys. Lett.* **88**, 023113 (2006).
24. Wrachtrup, J. & Jelezko, F. Processing quantum information in diamond. *J. Phys. Condens. Matter* **18**, S807–S824 (2006).
25. Chang, D. E., Sorensen, A. S., Hemmer, P. R. & Lukin, M. D. Quantum optics with surface plasmons. *Phys. Rev. Lett.* **97**, 053002 (2006).
26. Balasubramanian, G. *et al.* Nanoscale imaging magnetometry with diamond spins under ambient conditions. *Nature* doi:10.1038/nature07278 (this issue).
27. Degen, C. L. Scanning magnetic field microscope with a diamond single-spin sensor. Preprint at (<http://arXiv.org/abs/0805.1215v2>) (2008).
28. Cirac, J. I. & Zoller, P. A scalable quantum computer with ions in an array of microtraps. *Nature* **404**, 579–581 (2000).
29. Treutlein, P., Hunger, D., Camerer, S., Hansch, T. W. & Reichel, J. Bose-Einstein condensate coupled to a nanomechanical resonator on an atom chip. *Phys. Rev. Lett.* **99**, 140403 (2007).
30. Rabl, P. *et al.* Strong magnetic coupling between an electronic spin qubit and a mechanical resonator. Preprint at (<http://arXiv.org/abs/0806.3606>) (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge A. Akimov, D. Budker, F. Jelezko, F. Koppens, A. Trifonov, P. Hemmer and J. Wrachtrup for many discussions and experimental assistance. This work was supported by the NSF, DARPA, the Packard Foundation and Harvard CNS.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.D.L. ([lukin@fas.harvard.edu](mailto:lukin@fas.harvard.edu)).

## METHODS

**Samples.** AC magnetometry was performed at room temperature on nitrogen-vacancy centres in both a bulk single-crystal diamond sample (1 mm × 1 mm × 0.5 mm, natural diamond with an atypically low nitrogen concentration) and in diamond nanocrystals (monocrystalline, synthetic diamonds, 30 nm mean diameter, purchased from Microdiamant) deposited on a quartz coverslip.

**Confocal set-up.** Single nitrogen-vacancy centres were isolated and probed via confocal microscopy. Phonon-mediated fluorescent emission (630–750 nm) was detected under coherent optical excitation ( $\lambda = 532$  nm) using a single photon counting module (Perkin-Elmer SPCM-AQRH-13). The density of nitrogen-vacancy centres in both the bulk single-crystal and nanocrystal samples were sufficiently low that single bright spots (within the approximate confocal volume of 200 nm × 200 nm × 500 nm) were resolvable from the background fluorescence. As single spots in the confocal image may constitute many nitrogen-vacancy centres, single centres were identified by observing photon antibunching in the measurement of the second-order correlation function. This emission was separated from the excitation path using a dichroic mirror, and also notch and longpass filters. Samples were imaged with an oil immersion objective lens (Nikon CFI Plan Fluor Series, NA = 1.3, 100× magnification) over a 50  $\mu$ m × 50  $\mu$ m area in the plane normal to the optical path. Two galvanometer controlled mirrors steered the beam path for rapid imaging of this area. Experimental drifting of the focal plane due to thermal effects was compensated by using closed-loop feedback of the galvanometer and objective piezo voltages.

**Single-centre electron spin resonance.** The nitrogen-vacancy centre  $^3A_2$  ground state consists of two unpaired electrons in a triplet configuration leading to a zero-field splitting ( $\Delta = 2.87$  GHz) between the  $m_s = 0$  and  $m_s = \pm 1$  sublevels. Coherent optical excitation at  $\lambda = 532$  nm optically pumped the ground state into its  $m_s = 0$  sublevel. In addition, an external static magnetic field produced by a pair of Helmholtz coils split the degeneracy between the  $m_s = \pm 1$  states. It was then possible to selectively address transitions between the  $m_s = 0$  and  $m_s = 1$  (or  $m_s = -1$ ) states with microwave radiation (Fig. 1b) and manipulate a two-level subspace of the spin triplet (for example with spin-echo pulse sequences). Microwave radiation was applied by using the magnetic field emanating from a 20  $\mu$ m wire placed on the surface of the samples.

The excited ( $^3E$ ) state decay rates, also responsible for optical pumping, provided a means for spin-sensitive detection, as the rate of fluorescence was reduced for the  $m_s = \pm 1$  states compared to the  $m_s = 0$  states, with >35% contrast. The spin state in the ground electronic state was measured by pulsing on green excitation and monitoring the total number of photons collected within the optimal measurement interval, 324 ns. A 300 MHz PulseBlaster ESR pulse generator was employed for timing the triggering of counters, microwave pulses, the AC magnetic field, and the excitation laser. Microwave pulses were provided by gating the output of a frequency synthesizer with a microwave switch, while green laser pulses were generated using an acousto-optic modulator. The  $\pi$  and

$\pi/2$  pulses used for the spin-echo sequence were calibrated from the Rabi nutation curves between the two spin states.

**AC magnetometry.** As described in Fig. 1d and in ref. 8, demonstration magnetometer measurements were performed for an externally applied AC magnetic field with amplitude  $B_{AC}$ , frequency  $\nu$ , and phase  $\phi_{AC}$  during a cycle of a spin-echo sequence with a period  $\tau$ . The accumulated phase of the spin superposition state

$$\delta\phi = \frac{4g\mu_B B_{AC}}{2\pi\nu} \sin^2\left(\frac{\pi\nu\tau}{2}\right) \cos(\pi\nu\tau + \phi_{AC}) \quad (2)$$

contained information about the projection of the AC magnetic field amplitude onto the quantization axis of the electronic spin, corresponding to a vector magnetometer. Oscillatory magnetic fields from 1–10 kHz were generated by modulating the current through a Helmholtz pair also used to apply a bias DC magnetic field. For application of higher frequency AC fields (100–300 kHz), a single coil (60 turns) was resonantly driven and placed near the sample.

The measured signal intensity  $S_B$  was a function of the accumulated phase  $\delta\phi$ , as given by the probability of being in the  $m_s = 0$  state after the spin-echo pulse sequence:  $S_B \propto P_0(B_{AC}) = [1 + F(\tau)\cos(\delta\phi)]/2$ . Ideally, for a single-shot measurement of  $B_{AC}$  the sensitivity was maximized for a particular  $\nu$  by setting  $\tau = 1/\nu$ . In practice, many spin-echo cycles were averaged to reduce the uncertainty in photon statistics given the low single-shot count rate. To this end, the period of the entire measurement sequence (including polarization and readout, Fig. 1d) was matched to  $1/\nu$  in order to avoid multiple offset phases  $\phi_{AC}$  when the periods were incommensurate. The dependence on  $\phi_{AC}$  was removed entirely by appropriately shifting the time origin of the measurement pulse train. As the polarization ( $\tau_p = 1$   $\mu$ s) and readout ( $\tau_r = 3$   $\mu$ s) periods were short compared to the oscillation periods for typical 1–10 kHz AC magnetic fields, this choice introduced a slight deviation  $\varepsilon$  from the optimal  $\delta\phi$ , as  $\tau_p + \tau_r + \tau = 1/\nu \rightarrow \tau\nu = 1 - \varepsilon$ . The overall sensitivity was thus slightly reduced from its optimal value as  $1 - \mathcal{O}(\varepsilon^2)$ . For all experiments presented here,  $\nu\tau = 0.88$  was used. The envelope of the spin-echo signal,  $F(\tau)$  (see, for example, Fig. 2a) was modelled with an exponential decay modulated by the effect of a pair of nearby strongly interacting  $^{13}\text{C}$  nuclear spins. In this model<sup>21</sup>,  $F(\tau) = \exp(-(\tau/T_2)^4)(1 - [(a^2 - b^2)/a^2]\sin^2(a\tau)\sin^2(b\tau))$ ; where for the data in Fig. 2a we found  $T_2 = 676$   $\mu$ s,  $b = 478$  Hz (corresponding to the dipolar interaction between the two nuclei) and  $a = 626$  Hz (corresponding to the interactions between the nuclei and the nitrogen-vacancy spin). Using these experimentally determined parameters, the above model provided a prediction for the magnetometer sensitivity<sup>8</sup>  $\eta_{AC} = \pi\hbar/(g\mu_B C\sqrt{\nu}F(1/\nu))$  as a function of frequency (solid curve in Fig. 3a), where  $g \approx 2$  is the electron g-factor,  $\mu_B$  is the Bohr magneton, and  $C^{-2} = 1 + 2(a_0 + a_1 + a_0a_1)/(a_0 - a_1)^2$  is a factor that estimates<sup>8</sup> the photon shot noise when the average photon number during the readout window of 324 ns is much less than 1. The values  $a_0 = 0.03 \pm 0.006$  and  $a_1 = 0.018 \pm 0.004$  were the average numbers of detected photons for the electronic spin states  $m_s = 0$  and  $m_s = \pm 1$ , respectively.



## LETTERS

# Nanoscale imaging magnetometry with diamond spins under ambient conditions

Gopalakrishnan Balasubramanian<sup>1</sup>, I. Y. Chan<sup>2†</sup>, Roman Kolesov<sup>1</sup>, Mohannad Al-Hmoud<sup>1</sup>, Julia Tisler<sup>1</sup>, Chang Shin<sup>3</sup>, Changdong Kim<sup>3</sup>, Aleksander Wojcik<sup>3</sup>, Philip R. Hemmer<sup>3</sup>, Anke Krueger<sup>4</sup>, Tobias Hanke<sup>5</sup>, Alfred Leitenstorfer<sup>5</sup>, Rudolf Bratschitsch<sup>5</sup>, Fedor Jelezko<sup>1</sup> & Jörg Wrachtrup<sup>1</sup>

Magnetic resonance imaging and optical microscopy are key technologies in the life sciences. For microbiological studies, especially of the inner workings of single cells, optical microscopy is normally used because it easily achieves resolution close to the optical wavelength. But in conventional microscopy, diffraction limits the resolution to about half the wavelength. Recently, it was shown that this limit can be partly overcome by nonlinear imaging techniques<sup>1,2</sup>, but there is still a barrier to reaching the molecular scale. In contrast, in magnetic resonance imaging the spatial resolution is not determined by diffraction; rather, it is limited by magnetic field sensitivity, and so can in principle go well below the optical wavelength. The sensitivity of magnetic resonance imaging has recently been improved enough to image single cells<sup>3,4</sup>, and magnetic resonance force microscopy<sup>5</sup> has succeeded in detecting single electrons<sup>6</sup> and small nuclear spin ensembles<sup>7</sup>. However, this technique currently requires cryogenic temperatures, which limit most potential biological applications<sup>8</sup>. Alternatively, single-electron spin states can be detected optically<sup>9,10</sup>, even at room temperature in some systems<sup>11–14</sup>. Here we show how magneto-optical spin detection can be used to determine the location of a spin associated with a single nitrogen-vacancy centre in diamond with nanometre resolution under ambient conditions. By placing these nitrogen-vacancy spins in functionalized diamond nanocrystals, biologically specific magnetofluorescent spin markers can be produced. Significantly, we show that this nanometre-scale resolution can be achieved without any probes located closer than typical cell dimensions. Furthermore, we demonstrate the use of a single diamond spin as a scanning probe magnetometer to map nanoscale magnetic field variations. The potential impact of single-spin imaging at room temperature is far-reaching. It could lead to the capability to probe biologically relevant spins in living cells.

The nitrogen-vacancy centre in diamond is a unique solid state system that allows ultrasensitive and rapid detection of single electronic spin states under ambient conditions<sup>12</sup>. The nitrogen-vacancy defect is a naturally occurring impurity that is responsible for the pink colouration of diamond crystals when present in high concentration. It was demonstrated that this colour centre can be produced in diamond nanocrystals by electron irradiation. Fluorescing nitrogen-vacancy diamond nanocrystals can be used as markers for bioimaging applications<sup>15</sup>. Such markers have attracted widespread interest because of their unprecedented photostability and non-toxicity<sup>16,17</sup>. It was recognized recently that the magnetic properties of such fluorescent labels can in principle be used for novel microscopy<sup>18,19</sup>. Here we demonstrate the realization of a magneto-optic

microscope using nitrogen-vacancy diamond as the magnetic fluorescent label that moreover does not bleach or blink.

Figure 1c and d show the fluorescence and atomic force microscope image of nanocrystals containing nitrogen-vacancy defects. By careful choice of irradiation doses, we were able to control the number of nitrogen-vacancy centres per nanocrystal. The particular sample presented in Fig. 1 has on average a single nitrogen-vacancy defect per 40 nm nanocrystal (confirmed by fluorescence correlation measurements, Fig. 1e).

The energy level scheme and structure of the nitrogen-vacancy defect is shown in Fig. 1a and b. Two out of six electrons of the centre are unpaired, forming an electron spin triplet in the electronic ground and first excited state. Broadband optical excitation of the centre polarizes it by optical pumping into the  $m_s = 0$  spin sublevel. Laser-assisted detection of the spin state of a single nitrogen-vacancy centre makes use of differences in the absorption and emission properties of the spin sublevels. Specifically, the spin sublevel with magnetic quantum number  $m_s = 0$  (bright state) scatters  $\sim 30\%$  more photons than  $m_s = \pm 1$  states. Hence, when a resonant microwave field induces magnetic dipole transitions between these electronic spin sublevels, it destroys the optically pumped spin polarization, resulting in a significant decrease of the nitrogen-vacancy centre fluorescence. An example of such an optically detected electron spin resonance (ESR) spectrum of a single nitrogen-vacancy electronic spin is shown in Fig. 1f.

The spin Hamiltonian of the nitrogen-vacancy defect (neglecting electron–nuclear spin coupling) can be written as the sum of zero-field and Zeeman terms,  $H = D(S_z^2 - (1/3)[S(S+1)]) + E(S_x^2 - S_y^2) + g\mu_B \mathbf{B} \cdot \mathbf{S}$ , where  $D$  and  $E$  are zero-field splitting parameters,  $S = 1$ ,  $\mu_B$  is the Bohr magneton and  $g$  is the electron  $g$ -factor ( $g = 2.0$ ). Owing to the magnetic dipole interaction between the two unpaired electrons even at zero external magnetic field, the sublevels  $m_s = 0$  and  $m_s = \pm 1$  are separated ( $D = 2,870$  MHz). Owing to symmetry, the  $m_s = \pm 1$  sublevels of the nitrogen-vacancy defect are degenerate at zero magnetic field ( $E = 0$ ), resulting in a single resonance line appearing in the ESR spectrum (Fig. 1f). An external magnetic field lifts the degeneracy of  $m_s = \pm 1$ , leading to the appearance of two lines. By measuring the positions of the ESR resonances  $\omega_1$  and  $\omega_2$ , it is possible to calculate the magnitude of the external field  $B$  according to  $(g\mu_B)^2 = (1/3)(\omega_1^2 + \omega_2^2 - \omega_1\omega_2 - D^2) - E^2$  (see Methods for details).

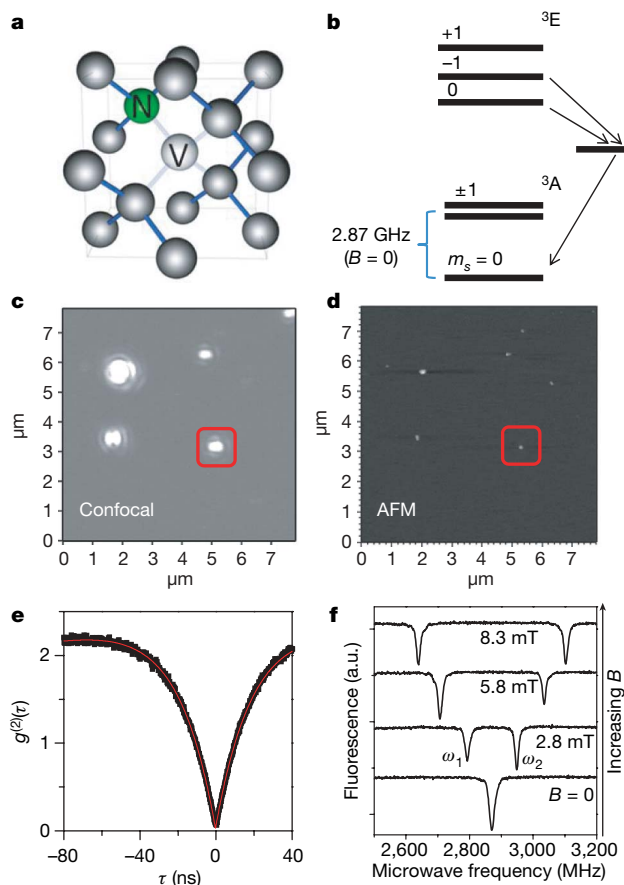
From the above-mentioned relations, it can be seen that, when combined with nano-positioning instrumentation, the single spin associated with a nitrogen-vacancy defect can be used as an atom-sized scanning probe vector magnetometer. Similarly, when placed in

<sup>1</sup>Physikalisches Institut, Universität Stuttgart, 70550 Stuttgart, Germany. <sup>2</sup>Department of Chemistry, Brandeis University, Waltham, Massachusetts 02454, USA. <sup>3</sup>Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas 77843, USA. <sup>4</sup>Otto-Diels-Institut für Organische Chemie, Christian-Albrechts-Universität zu Kiel, 24098 Kiel, Germany. <sup>5</sup>University of Konstanz and Center for Applied Photonics, 78457 Konstanz, Germany. <sup>†</sup>Present address: 3 Physikalisches Institut, Universität Stuttgart, 70550 Stuttgart, Germany.

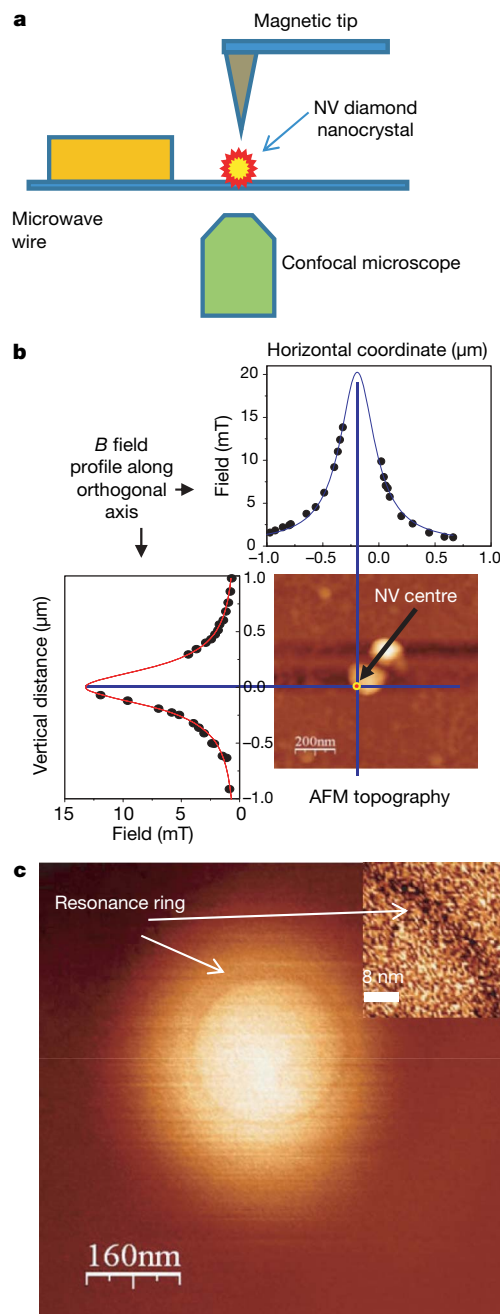
an inhomogeneous magnetic field with a known field gradient, the defect can be used as a magneto-optical spin marker for suboptical-wavelength tagged imaging. As a demonstration, two-dimensional spin imaging experiments were performed using a single nitrogen-vacancy centre and the highly inhomogeneous magnetic field produced by the magnetic tip of an atomic force microscope (AFM). The experimental set-up is shown in Fig. 2a. A commercial AFM was combined with a confocal microscope. The magnetic probe, commonly used in magnetic force microscopy, consists of a sharp silicon tip coated with 30 nm of magnetic material: the exact magnetic field profile of the cantilever is not known a priori, and must be determined. For this, we have used our single-spin nitrogen-vacancy magnetometer. The magnetic cantilever was first placed at a known distance from the diamond nanocrystal, and the magnetic field experienced by the single nitrogen-vacancy centre was recorded in steps (corresponding to several hundred nanometre displacements of the cantilever) by acquiring ESR spectra such as those in Fig. 1f at each location. The experimentally obtained data points were then fitted using a Lorentzian function, inferred from numerical simulation of the field created by the cantilever (Fig. 2b). This gives the magnetic field profile of the cantilever in one dimension. Similarly, the profile along an orthogonal axis is recorded to give the

two-dimensional profile as well as the exact position of the nitrogen-vacancy centre.

To visualize the resolving power of our gradient imaging technique, the magnetic cantilever was scanned in the vicinity of a nanocrystal containing a single nitrogen-vacancy defect while simultaneously exciting with a fixed-frequency microwave field. When a confocal image is acquired, each point of the optical image corresponds to a well-defined magnetic field value (as measured in



**Figure 1 | Nitrogen-vacancy defect in diamond.** **a**, Structure and energy level scheme of the nitrogen-vacancy (NV) defect in diamond. Optical pumping initializes the centre into the  $m_s = 0$  spin state via spin selective shelving into the metastable singlet state,  $^1A$ . This state decays preferentially into the  $m_s = 0$  sublevel of the ground state, leading to optically induced spin polarization (more than >90% at room temperature). **c**, **d**, Simultaneously acquired optical (**c**) and AFM (**d**) image of diamond nanocrystals containing single nitrogen-vacancy defects. **e**, Fluorescence autocorrelation function, confirming that the nanocrystal contains a single nitrogen-vacancy defect. The contrast of  $g^2(\tau)$  at zero delay time scales as  $1/N$ , where  $N$  is the number of emitters. **f**, Optically detected magnetic resonance spectra for a single nitrogen-vacancy defect at increasing magnetic field (from bottom to top).



**Figure 2 | Gradient imaging with single spins.** **a**, Two-dimensional imaging is achieved using a field gradient created by a magnetic cantilever. **b**, The experimental profile of the cantilever's magnetic field for two orthogonal axes. The magnetic tip was placed at several points parallel to the blue lines, and the ESR spectra were measured. The calculated (fitted) magnetic field profile allows estimation of the location of the single nitrogen-vacancy centre (shown in the AFM topography). **c**, Two-dimensional magnetic resonance image of a single nitrogen-vacancy centre, showing resonance rings corresponding to a magnetic field of 3 mT (resonance frequency of 2,780 MHz). Inset, an enlarged section of a ring with a width of approximately 5 nm.

the previous experiment). At particular positions (pixels) when the microwave frequency is resonant with the corresponding spin sub-level splitting, the fluorescence intensity is reduced. This results in a dark ring (Fig. 2), which marks a two-dimensional cut through the  $B$  field corresponding to a constant magnetic field projection along the nitrogen-vacancy quantization axis. The width of the rings is given by the magnetic resonance linewidth divided by the field gradient ( $80 \mu\text{T nm}^{-1}$ ). This ring width also defines the ultimate resolution limit of this technique for spin imaging. Rings with 5 nm width are observed, and shown in Fig. 2c inset. Note that this width is smaller than the sizes of both the magnetic tip and the diamond nanocrystal, and is only possible because a single nitrogen-vacancy centre is localized to a fraction of a nanometre in the diamond lattice. The dark ring shown in Fig. 2c is only seen if simultaneously the nitrogen-vacancy axis is oriented vertically, and the magnetic field is radially symmetric. This special case was selected to simplify understanding of the technique. It is interesting to note that a vibrating cantilever (a.c. mode AFM was used in all the experiments) induces significant line broadening when the magnetic cantilever comes very close to the spin (see Supplementary Information).

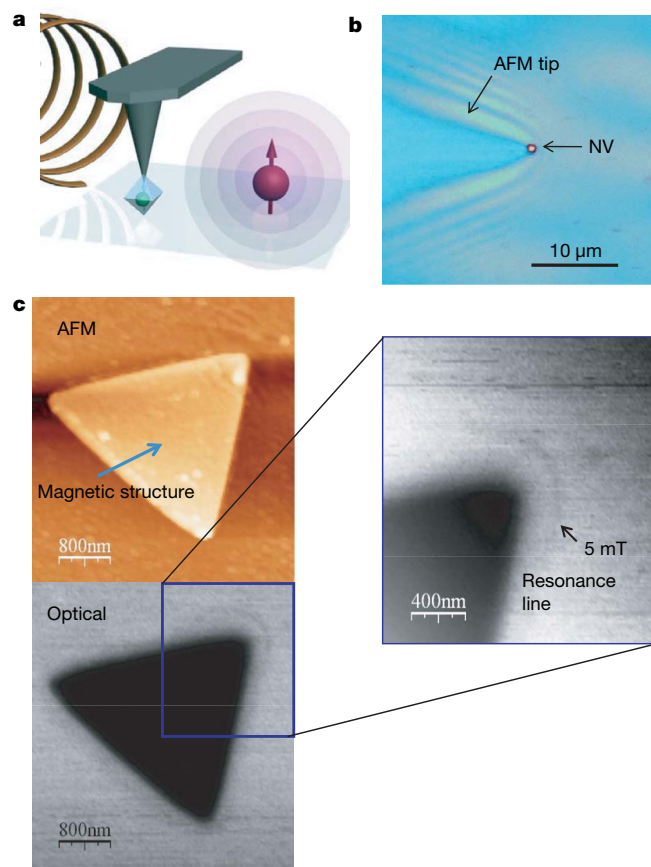
Being an atomic-sized non-perturbing magnetic field sensor, the single nitrogen-vacancy centre can be incorporated into the cantilever instead of a magnetic coating, and used as a scanning probe magnetometer to achieve subwavelength imaging resolution. To demonstrate the feasibility of this approach, we attached a nanocrystal containing a single nitrogen-vacancy centre to the tip of a cantilever, and used it to profile the magnetic field produced by a nanometre-sized magnetic structure. Details of the set-up are shown in Fig. 3a. Microwaves are tuned into resonance with the nitrogen-vacancy spin at the tip of the cantilever (see Fig. 3b) for a particular magnetic field projection. Hence the resonance conditions in the vicinity of the magnetic nanostructures are satisfied along well-defined lines of constant  $B_z$ , where  $z$  is along the nitrogen-vacancy quantization axis. Figure 3c shows a magneto-optical image of a triangular magnetic structure obtained with a single nitrogen-vacancy defect as light source (as expected, the structure appears as a shadow in our detection geometry). The narrow dark line close to the corner represents spatial regions where the conditions for magnetic resonance of the nitrogen-vacancy centre on the tip are fulfilled ( $B_z = 5 \text{ mT}$ ). Note that the image represents raw data acquired in just 4 minutes. The magnetic field resolution is given by the width of the dark lines, which are about 20 nm, multiplied by the magnetic field gradient of  $25 \mu\text{T nm}^{-1}$  (measured by recording several resonance lines at different microwave frequencies, data not shown). It corresponds to a measurement resolution of 0.5 mT. The limiting factor here is oscillatory motion of the nanodiamond attached to the AFM tip (see Methods for details).

The resolution could be significantly improved by phase locking of the detection system to the oscillatory motion of the cantilever, and using echo-based techniques with an echo period matched to a single oscillation period of the cantilever. This essentially corresponds to measuring a.c. instead of d.c. magnetic fields. The advantage of using echoes is that the effective ESR linewidth is narrower than the inhomogeneous linewidth<sup>20</sup>, and for a long spin phase memory time  $T_2$  is effectively given by the AFM vibration frequency, which is 100 kHz for standard AFM cantilevers. Hence we expect an improvement of field measurement accuracy by a factor of 150 ( $3 \mu\text{T}$ ) using this technique. For the magnetic field gradient caused by the structure imaged in Fig. 3c, this would correspond to subnanometre spatial resolution.

Ultrasensitive magnetometry with single spins in diamond not only allows high spatial resolution imaging, but also might be applied to image single external spins under ambient conditions. Here the magnetic sublevels of the nitrogen-vacancy centre are shifted by the magnetic fields produced by (for example) other single electron or nuclear magnetic dipoles in nearby molecules. To show the feasibility of single electron and nuclear spin detection, we estimate the ultimate sensitivity limit of a scanning spin microscope based on nitrogen-vacancy

centres in diamond. The magnetic field created by a single electron spin located at distance  $r$  from the nitrogen-vacancy spin is  $B^{\text{dip}} = (\mu_0 \mu / 4\pi) \sqrt{3 \cos^2 \theta + 1} / r^3$ , where  $\mu$  is the single spin magnetic moment, and  $\theta$  is the angle between the vector connecting the two spins and the vector of the external magnetic field. When we substitute  $\mu = -(1/2)g_e \mu_B \approx 10^{-23} \text{ JT}^{-1}$ , and  $\mu_0 / 4\pi \approx 10^{-7} \text{ NA}^{-2}$ , a field of  $10^{-5} \text{ T}$  can be obtained for a distance between the electron and nitrogen-vacancy spins of 5 nm. For the nitrogen-vacancy centre this gives up to 0.3 MHz of ESR frequency shift, which is within the projected detection limit for the single nitrogen-vacancy nanocrystals used in this demonstration. Imaging single nuclear spins is more challenging, as the lower nuclear magnetic moment results in fields three orders of magnitude lower ( $10^{-8} \text{ T}$ ). This value corresponds to a kilohertz shift of the nitrogen-vacancy resonance.

In general, the sensitivity of our nitrogen-vacancy magnetometer is determined by the linewidth of the ESR transition. Experiments presented here were carried out using continuous wave (c.w.) ESR



**Figure 3 | Scanning probe magnetometry.** **a**, Diagram of the magnetic field imaging experiment. A nanoscale magnetic particle (red) is imaged with a single nitrogen-vacancy defect (green, within the blue nanocrystal) fixed at the scanning probe tip (black). **b**, Optical image of a diamond nanocrystal attached to an AFM tip (view from the bottom). The scattered light image of the tip is overlapped with the fluorescence image of the nanocrystal. The bright spot (arrowed) represents fluorescence of a single nitrogen-vacancy defect. Fluorescence autocorrelation function (data not shown) shows a pronounced antibunching dip, indicating a single nitrogen-vacancy defect in the nanocrystal on the AFM tip. **c**, Field reconstruction using the scanning probe single spin magnetometer. Top left, an AFM image of a nickel magnetic nanostructure prepared by electron beam lithography; bottom left, a magneto-optical image of the same structure, recorded using a single nitrogen-vacancy centre on the AFM tip as light source and magnetometer. Inset (right), the fluorescence signal from the scanned nitrogen-vacancy centre attached to the apex of the AFM tip when resonant microwaves at 2,750 MHz are applied (the arrowed point corresponds to 5 mT resonance line with the magnetic field tilted by  $45^\circ$  relative to the nitrogen-vacancy axis).



and technical grade diamonds. Hence the linewidth of the spin resonance line (a few MHz) was limited by fast ( $\mu\text{s}$ ) decoherence and microwave-field-induced broadening. However, it was recently shown that the phase memory time for ultrapure diamond reaches one millisecond when echo-based techniques are used for detection<sup>21</sup>. This corresponds to a linewidth of the order of 0.3 kHz. Taking this value, single nuclear spins can be detected at 5 nm distance under ambient conditions<sup>22</sup>. As these spin linewidths were obtained under ambient conditions, this approach will potentially enable the use of nitrogen-vacancy defects in diamond nanocrystals as a probe for intracellular electron (and possibly nuclear) spin imaging.

## METHODS SUMMARY

Magnetic imaging and magnetometry experiments were performed using a home-built scanning confocal microscope combined with an AFM (MFP-3D Asylum Research). Nitrogen-vacancy defects were excited with a frequency doubled c.w. Nd:YAG laser (Coherent Compass) focused on to the sample with a high NA objective (Olympus PlanAPO, NA = 1.35). Luminescence light was collected by the same objective and filtered from the excitation light using a dichroic beamsplitter (640 DCXR, Chroma) and long-pass filter (647 LP, Chroma). Photon counting of the filtered light was performed using two avalanche photodiodes (SPQR-14, Perkin-Elmer). Fluorescence autocorrelation histograms were recorded using a fast multichannel analyser (Fastcomtec, P7889). Optically detected magnetic resonance measurements were performed using a commercial microwave source (Rhode & Schwarz GmbH, SMIQ 03) amplified by a travelling wave tube amplifier (Hughes 8020H). Commercially available magnetic cantilevers (Team Nanotec) were used for generation of high magnetic field gradients. UV curing glue (Thorlabs) was used to attach diamond nanocrystals to the AFM tip for the scanning probe magnetometry.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 23 April; accepted 18 July 2008.

- Willig, K. I., Rizzoli, S. O., Westphal, V., Jahn, R. & Hell, S. W. STED microscopy reveals that synaptotagmin remains clustered after synaptic vesicle exocytosis. *Nature* **440**, 935–939 (2006).
- Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
- Aguayo, J. B., Blackband, S. J., Schoeniger, J., Mattingly, M. A. & Hintermann, M. Nuclear-magnetic-resonance imaging of a single cell. *Nature* **322**, 190–191 (1986).
- Ciobanu, L., Seeber, D. A. & Pennington, C. H. 3D MR microscopy with resolution 3.7  $\mu\text{m}$  by 3.3  $\mu\text{m}$  by 3.3  $\mu\text{m}$ . *J. Magn. Reson.* **158**, 178–182 (2002).
- Rugar, D., Yannoni, C. S. & Sidles, J. A. Mechanical detection of magnetic resonance. *Nature* **360**, 563–566 (1992).
- Rugar, D., Budakian, R., Mamin, H. J. & Chui, B. W. Single spin detection by magnetic resonance force microscopy. *Nature* **430**, 329–332 (2004).

- Mamin, H. J., Poggio, M., Degen, C. L. & Rugar, D. Nuclear magnetic resonance imaging with 90-nm resolution. *Nature Nanotechnol.* **2**, 301–306 (2007).
- Glover, P. & Mansfield, P. Limits to magnetic resonance microscopy. *Rep. Prog. Phys.* **65**, 1489–1511 (2002).
- Kohler, J. *et al.* Magnetic resonance of a single molecular spin. *Nature* **363**, 242–244 (1993).
- Wrachtrup, J., von Borczyskowski, C., Bernard, J., Orrit, M. & Brown, R. Optical detection of magnetic resonance in a single molecule. *Nature* **363**, 244–245 (1993).
- Hanson, R., Dobrovitski, V. V., Feiguin, A. E., Gywat, O. & Awschalom, D. D. Coherent dynamics of a single spin interacting with an adjustable spin bath. *Science* **320**, 352–355 (2008).
- Gruber, A. *et al.* Scanning confocal optical microscopy and magnetic resonance on single defect centers. *Science* **276**, 1202–1204 (1997).
- Epstein, R. J., Mendoza, F. M., Kato, Y. K. & Awschalom, D. D. Anisotropic interactions of a single spin and dark-spin spectroscopy in diamond. *Nature Phys.* **1**, 94–98 (2005).
- Childress, L. *et al.* Coherent dynamics of coupled electron and nuclear spin qubits in diamond. *Science* **314**, 281–285 (2006).
- Fu, C. C. *et al.* Characterization and application of single fluorescent nanodiamonds as cellular biomarkers. *Proc. Natl Acad. Sci. USA* **104**, 727–732 (2007).
- Liu, K. K., Cheng, C. L., Chang, C. C. & Chao, J. I. Biocompatible and detectable carboxylated nanodiamond on human cell. *Nanotechnology* **18**, 325102 (2007).
- Neugart, F. *et al.* Dynamics of diamond nanoparticles in solution and cells. *Nano Lett.* **7**, 3588–3591 (2007).
- Chernobrod, B. M. & Berman, G. P. Spin microscope based on optically detected magnetic resonance. *J. Appl. Phys.* **97**, 014903 (2005).
- Kuhn, S., Hettich, C., Schmitt, C., Poizat, J. P. H. & Sandoghdar, V. Diamond colour centres as a nanoscopic light source for scanning near-field optical microscopy. *J. Microsc.* **202**, 2–6 (2001).
- Taylor, J. M. *et al.* High-sensitivity diamond magnetometer with nanoscale resolution. *Nature Phys.* (in the press); preprint at (<http://arXiv.org/abs/0805.1367v1>) (2008).
- Gaebel, T. *et al.* Room-temperature coherent coupling of single spins in diamond. *Nature Phys.* **2**, 408–413 (2006).
- Maze, J. R. *et al.* Nanoscale magnetic sensing with an individual electronic spin in diamond. *Nature* doi:10.1038/nature07279 (this issue).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. D. Lukin for drawing our attention to advanced echo-based techniques, and R. Kamella for technical assistance. This work was supported by the EU (QAP, EQUIND, NANO4DRUGS, NEDQIT), DFG (SFB/TR21 and FOR730) and Landesstiftung BW.

**Author Contributions** G.B., I.Y.C., R.K., M.A.-H., J.T., C.S., C.K., A.W., J.W. and F.J. performed the experiments; A.K. prepared diamond nanocrystals; T.H., A.L. and R.B. prepared magnetic nanostructures; P.R.H., J.W. and F.J. designed and coordinated the experiments; and F.J. wrote the paper. All authors discussed the results, analysed the data and commented on the manuscript.

**Author Information** Correspondence and requests for materials should be addressed to F.J. ([f.jelezko@physik.uni-stuttgart.de](mailto:f.jelezko@physik.uni-stuttgart.de)).

## METHODS

### Vector magnetometry using a single nitrogen-vacancy defect electron spin.

The resonance frequencies of the  $m_s = 0 \leftrightarrow \pm 1$  spin transitions of a spin-1 system allow one to extract the magnitude of the local magnetic field and, under some approximations, the angle between the magnetic field and the symmetry axis of the system. The spin Hamiltonian of an  $S = 1$  system having a distorted  $C_{3v}$  symmetry is given by the following expression:

$$H = \mu_B g \mathbf{B} \cdot \mathbf{S} + D(S_z^2 - S(S+1)/3) + E(S_x^2 - S_y^2)$$

where  $D$  and  $E$  are the zero-field splitting parameters,  $S = 1$ ,  $\mu_B$  is the Bohr magneton,  $g = 2$  is the  $g$ -factor, and  $\mathbf{B}$  is the external magnetic field. Imperfect axial symmetry is reflected by the asymmetry parameter  $E$ . These parameters unambiguously determine the natural local coordinate system, with the  $z$  axis being along the axis of the nitrogen-vacancy centre and  $x$  and  $y$  axes being along the principal axes of the distortion ellipsoid. In such a coordinate system, it is convenient to describe the magnetic field by its magnitude  $B$  and the two angles,  $\theta$  (polar) and  $\varphi$  (azimuthal). All parameters  $B$ ,  $\theta$  and  $\varphi$  can be obtained from analysis of the ESR spectrum. The positions of spin levels are given by the solutions of the characteristic equation:

$$x^3 - \left(\frac{D^2}{3} + E^2 + \beta^2\right)x - \frac{\beta^2}{2}(D\cos 2\theta + 2E\cos 2\varphi \sin^2 \theta) - \frac{D}{6}(4E^2 + \beta^2) + \frac{2D^3}{27} = 0$$

where  $\beta = \mu_B g B$ . Denoting the frequency of the  $S_z = 0$  state as  $x_0$ , one finds that the positions of the  $S_z = \pm 1$  states are given by  $x_{\pm} = x_0 + v_{0\pm}$ , where  $v_{0\pm}$  are the experimentally measured frequencies of  $0 \leftrightarrow \pm 1$  spin transitions. Since  $x_{\pm}$  must satisfy the above-mentioned equation, it is possible to obtain three equations for the four unknowns ( $x_0$ ,  $B$ ,  $\theta$  and  $\varphi$ ), two of which,  $\theta$  and  $\varphi$ , form a unique combination  $A = D\cos 2\theta + 2E\cos 2\varphi \sin^2 \theta$ . This set of equations gives the following solutions for  $\beta$  and  $A$ :

$$\beta^2 = \frac{1}{3}(v_1^2 + v_2^2 - v_1 v_2 - D^2) - E^2,$$

$$A = \frac{7D^3 + 2(v_1 + v_2)(2(v_1^2 + v_2^2) - 5v_1 v_2 - 9E^2) - 3D(v_1^2 + v_2^2 - v_1 v_2 + 9E^2)}{9(v_1^2 + v_2^2 - v_1 v_2 - D^2 - 3E^2)}$$

Since for nitrogen-vacancy centres  $D \gg E$ ,  $A \approx D\cos 2\theta$ . Thus, knowing the zero-field splitting parameters and the frequencies of the  $0 \leftrightarrow \pm 1$  ESR resonances, one can find  $B$  and  $\theta$ . The solution of the inverse problem of finding the two ESR frequencies given the known  $B$  and  $\theta$  is presented in Supplementary Information.

**Modelling the magnetic field of the cantilever.** The magnetic field of a tip having a ferromagnetic coating was simulated in the following manner. The surface of the tip is assumed to be an axially symmetric cone with a round apex. It can be simulated by the following simple analytical formula:

$$h = r \tanh \frac{r}{2r_0}$$

where  $h$  is the height of the surface point above the apex,  $r$  is the radial coordinate, and  $r_0$  is the curvature radius of the rounded apex of the tip. It is assumed that the tip surface is covered with a thin layer of ferromagnetic material. The magnetic field produced by an infinitely small element of the surface was approximated as that of a magnetic dipole. The magnetization of the surface is assumed to be uniform and the direction of the magnetization of each surface element assumed the same. The contributions of all surface elements were then integrated over the surface of the tip. We are interested in the magnetic field in a plane somewhat below the tip apex and perpendicular to the tip axis. In the simplest case of the magnetization pointing along the tip axis, the magnetic field has only radial and axial components. The typical result of a simulation is shown in Supplementary Information. It justifies the Lorentzian field distribution model used to find the two-dimensional position of the nitrogen-vacancy centre.

## LETTERS

# Thresholds for Cenozoic bipolar glaciation

Robert M. DeConto<sup>1</sup>, David Pollard<sup>2</sup>, Paul A. Wilson<sup>3</sup>, Heiko Pälike<sup>3</sup>, Caroline H. Lear<sup>4</sup> & Mark Pagani<sup>5</sup>

The long-standing view of Earth's Cenozoic glacial history calls for the first continental-scale glaciation of Antarctica in the earliest Oligocene epoch (~33.6 million years ago<sup>1</sup>), followed by the onset of northern-hemispheric glacial cycles in the late Pliocene epoch, about 31 million years later<sup>2</sup>. The pivotal early Oligocene event is characterized by a rapid shift of 1.5 parts per thousand in deep-sea benthic oxygen-isotope values<sup>3</sup> (Oi-1) within a few hundred thousand years<sup>4</sup>, reflecting a combination of terrestrial ice growth and deep-sea cooling. The apparent absence of contemporaneous cooling in deep-sea Mg/Ca records<sup>4–6</sup>, however, has been argued to reflect the growth of more ice than can be accommodated on Antarctica; this, combined with new evidence of continental cooling<sup>7</sup> and ice-rafted debris<sup>8,9</sup> in the Northern Hemisphere during this period, raises the possibility that Oi-1 represents a precursory bipolar glaciation. Here we test this hypothesis using an isotope-capable global climate/ice-sheet model that accommodates both the long-term decline of Cenozoic atmospheric CO<sub>2</sub> levels<sup>10,11</sup> and the effects of orbital forcing<sup>12</sup>. We show that the CO<sub>2</sub> threshold below which glaciation occurs in the Northern Hemisphere (~280 p.p.m.v.) is much lower than that for Antarctica (~750 p.p.m.v.). Therefore, the growth of ice sheets in the Northern Hemisphere immediately following Antarctic glaciation would have required rapid CO<sub>2</sub> drawdown within the Oi-1 time-frame, to levels lower than those estimated by geochemical proxies<sup>10,11</sup> and carbon-cycle models<sup>13,14</sup>. Instead of bipolar glaciation, we find that Oi-1 is best explained by Antarctic glaciation alone, combined with deep-sea cooling of up to 4 °C and Antarctic ice that is less isotopically depleted (–30 to –35‰) than previously suggested<sup>15,16</sup>. Proxy CO<sub>2</sub> estimates remain above our model's northern-hemispheric glaciation threshold of ~280 p.p.m.v. until ~25 Myr ago, but have been near or below that level ever since<sup>10,11</sup>. This implies that episodic northern-hemispheric ice sheets have been possible some 20 million years earlier than currently assumed (although still much later than Oi-1) and could explain some of the variability in Miocene sea-level records<sup>17,18</sup>.

Evidence for the onset of Antarctic glaciation comes from a combination of marine geochemical and sea-level records<sup>3,19,20</sup>, and more direct records of ice-rafted debris and glaciomarine sediments from around the Antarctic margin<sup>1</sup>. Proposed mechanisms include the opening of Southern Ocean gateways<sup>21</sup>, mountain uplift and orbital forcing<sup>22</sup>; however, recent modelling studies implicate low atmospheric CO<sub>2</sub> as the most important factor<sup>22</sup>. The growth of ice sheets in the Northern Hemisphere is thought to have begun much later than in Antarctica, beginning on southern Greenland in the late Miocene or early Pliocene<sup>23</sup> and culminating with the onset of major glacial cycles around 2.7–3.0 Myr ago<sup>3</sup>. Atmospheric CO<sub>2</sub> is also considered a critical factor for Northern Hemispheric glaciation, perhaps with additional influence from ocean gateways and mountain uplift<sup>24</sup>. The recent discovery of much older Eocene, Oligocene and Miocene ice-rafted debris in the Greenland Sea<sup>8,9</sup> and Arctic Ocean<sup>25</sup>

has called this long-standing view of Earth history into question, although the amount of Northern Hemispheric ice responsible for these sediments remains controversial.

Our current understanding of cryospheric evolution comes largely from marine oxygen isotope<sup>26</sup> and Mg/Ca records<sup>5,6</sup> from foraminiferal calcite, and stratigraphic sea-level records from passive continental margins<sup>19,20</sup>. Benthic isotope records reflect the combined effects of ice volume and ocean temperature, and in principle Mg/Ca ratios provide an independent record of temperature. These data can be combined to deconvolve the ice-volume component of the isotope records, which show a number of sudden and large (>1‰) shifts and excursions throughout the Cenozoic (for example, the Oi and Mi events)<sup>3</sup>. These shifts are thought to represent Antarctic glaciation events because direct evidence for significant northern-hemispheric ice is lacking before ~3.0 Myr ago. The stepwise shift in benthic  $\delta^{18}\text{O}$  in the earliest Oligocene (Oi-1) is the largest in the Cenozoic. As shown in a highly resolved record from Ocean Drilling Program site 1218 in the eastern tropical Pacific<sup>4</sup>, the event began around 34 Myr ago, during a minimum in the 1.2-Myr obliquity cycle producing a long interval of low seasonality (Fig. 1). This observation is in good agreement with time-continuous simulations from a combined general circulation model (GCM)/ice-sheet model accounting for gradual Cenozoic CO<sub>2</sub> decline and idealized orbital forcing<sup>22</sup>. In these simulations an Antarctic ice sheet grows suddenly once CO<sub>2</sub> reaches a critical threshold value around 2.8 to 2.6 times the 'pre-industrial' atmospheric level (PAL, taken to be 280 p.p.m.v., attaining a volume ( $21 \times 10^6 \text{ km}^3$ ) comparable to the modern East Antarctic ice sheet. The simulated ice sheet expands in two rapid jumps in response to height/mass-balance and albedo feedbacks initiated when snowlines intersect high plateaux during orbital periods producing cold austral summers<sup>22</sup>. The simulation bears a striking resemblance to the form of the site 1218 record<sup>4</sup>, which shows a two-step shift in oxygen isotopes separated by ~200 kyr, with each step occurring within one 41-kyr obliquity cycle (Fig. 1).

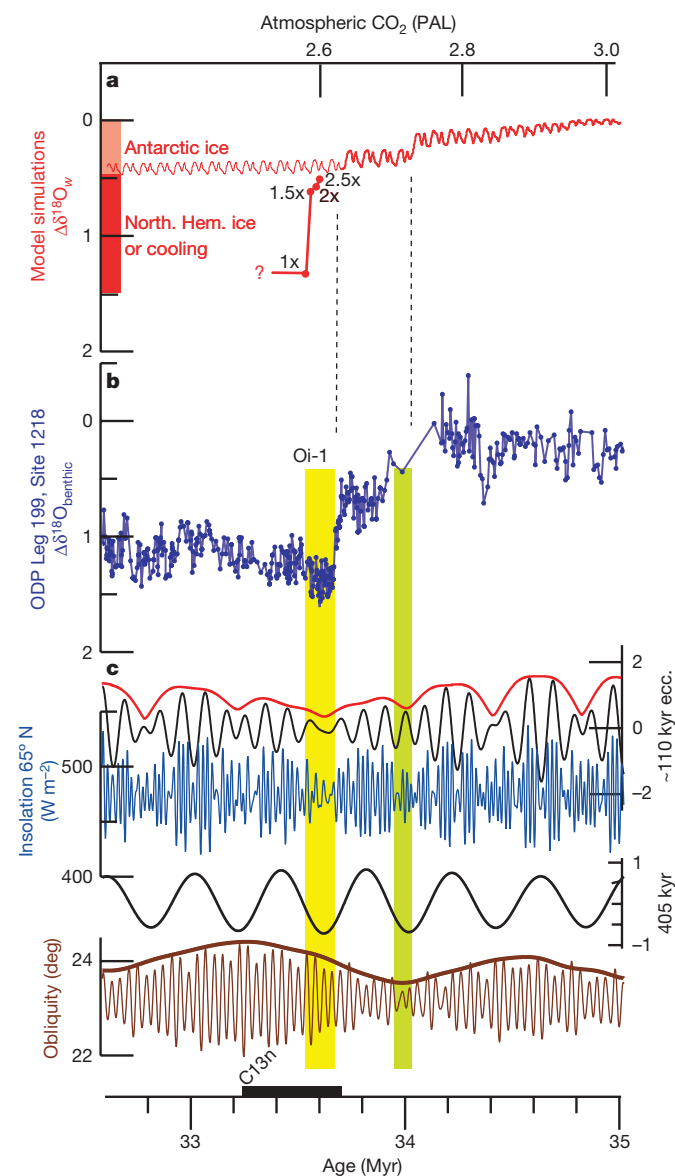
The problem lies in the magnitude of the isotope shift, which, if taken as the total change in benthic  $\delta^{18}\text{O}$  from the latest Eocene to the peak of Oi-1, is ~1.5‰. Assuming that the isotopic composition of Palaeogene Antarctic ice was similar to today (–45 to –57‰ for West and East Antarctica, respectively<sup>27</sup>), and ignoring changes in deep-sea temperature, the implied increase in ice volume is  $\sim 40 \times 10^6 \text{ km}^3$  or ~100 m of equivalent sea level. This is 135% of modern Antarctic ice volume and nearly twice the volume of the simulated Oi-1 ice sheet<sup>22</sup>. If the isotopic composition of precipitation falling on a warmer and thinner Oligocene ice sheet were less depleted than today as suggested previously<sup>4,22</sup>, the missing ice-volume problem would be greatly exacerbated.

To test the hypothesis that ancient ice on a warmer Antarctica was isotopically less depleted than today, we ran a set of oxygen-isotope simulations spanning the Eocene/Oligocene climate transition, using an isotope tracer-capable version<sup>28</sup> of the same GCM used in our prior

<sup>1</sup>Department of Geosciences, University of Massachusetts, Amherst, Massachusetts 01003, USA. <sup>2</sup>Earth and Environmental Systems Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA. <sup>3</sup>National Oceanography Centre, University of Southampton, Southampton SO14 3ZH, UK. <sup>4</sup>School of Earth and Ocean Sciences, Cardiff University, Cardiff CF10 3YE, UK. <sup>5</sup>Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520, USA.



simulations of Oi-1 (ref. 22; see Methods). Our results indicate that the first ice to accumulate in the earliest Oligocene would have had an isotopic composition of  $-20\text{‰}$  to  $-25\text{‰}$  (SMOW), becoming more



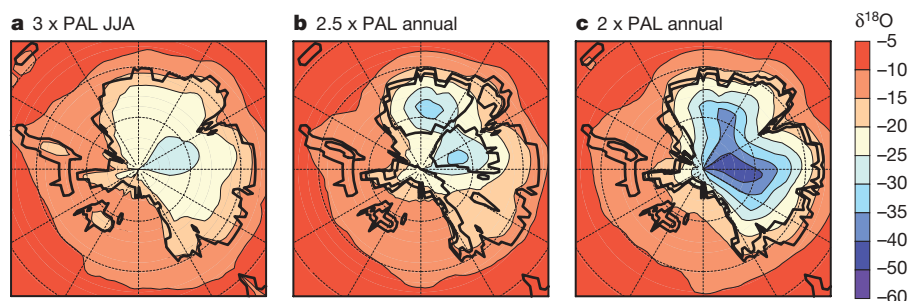
**Figure 1 | Changes in the isotopic composition of the ocean across the Eocene/Oligocene transition.** Isotopic, orbital and model time series are shown on the same astronomically tuned timescale<sup>4</sup>, with the simulated and observed stepwise timing of glaciation aligned (dashed lines) for comparison. **a**, The simulated change in mean ocean  $\delta^{18}\text{O}_w$  ( $\delta^{18}\text{O} = [(^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{standard}}] - 1$ , where standard is SMOW) from coupled GCM/ice-sheet simulations assuming the isotopic composition of ice is  $-35\text{‰}$ . Red bars at left show the relative contributions from Antarctic ice (light red) and from deep-sea cooling and/or Northern Hemisphere ice (dark red). The thin red line shows a prior simulation<sup>22</sup> ignoring Northern Hemisphere ice and assuming a continuous decline in  $\text{CO}_2$ . Carbon dioxide levels (top x axis) are the average of two simulations assuming open and closed Southern Ocean gateways<sup>22</sup>. A second scheme accounting for Northern Hemisphere ice sheets and rapidly decreasing  $\text{CO}_2$  beginning after Antarctic glaciation is shown by the thick red line, with red dots corresponding to the added effect of the ice shown in Fig. 3. **b**, High-resolution benthic  $\delta^{18}\text{O}$  data from Ocean Drilling Program site 1218 (ref. 4). **c**, Orbital parameters<sup>12</sup> include filtered and normalized eccentricity values for 110-kyr and 405-kyr periodicities (black lines) and the long-term envelope for the 110 kyr (red line). The initial step in Antarctic ice growth corresponds to low obliquity variance (green bar) and the second main step occurs during an interval of reduced summer insolation at high latitudes (yellow bar) as assumed in our Northern Hemisphere simulations.

negative as the ice sheet grew, but not exceeding  $-42\text{‰}$ , even on the highest ice elevations (Fig. 2). The average isotopic composition of precipitation in the accumulation zone of the fully developed earliest Oligocene ice sheet is  $-30$  to  $-35\text{‰}$ . In the absence of deep-sea cooling, Oi-1 would require the growth of  $\sim 65 \times 10^6 \text{ km}^3$  of ice (160 m equivalent sea level), far in excess of the holding capacity of a warmer Antarctic continent (see Methods and Supplementary Information).

To test the possibility that Oi-1 includes a contribution from northern-hemispheric ice sheets, we ran a series of 30-kyr simulations using a GCM/ice-sheet model<sup>18,22</sup>, developed specifically for simulating the time-continuous evolution of climate and ice sheets over long timescales. The simulations used an earliest Oligocene palaeogeography, with an Antarctic ice sheet already in place<sup>22</sup>. Orbital values were initialized with a favourable but reasonable configuration for northern-hemispheric glaciation, with some experiments updated every 10 kyr to account for subsequent orbital variations (see Methods). The initial cold boreal summer orbit (well within the range that occur on Cenozoic timescales<sup>12</sup>) was chosen to constrain the highest level of  $\text{CO}_2$  that would allow Northern Hemisphere glaciation, given the position and topography of Oligocene continents. Atmospheric  $\text{CO}_2$  was lowered by  $0.5 \times \text{PAL}$  in each successive simulation beginning with a starting value of  $2.5 \times \text{PAL}$ , just below the model's Antarctic  $\text{CO}_2$ -glaciation threshold. To substantiate our results in light of poorly constrained geographical boundary conditions, the entire sequence was repeated with most Northern Hemisphere continental elevations reduced by 50% (see Supplementary Information).

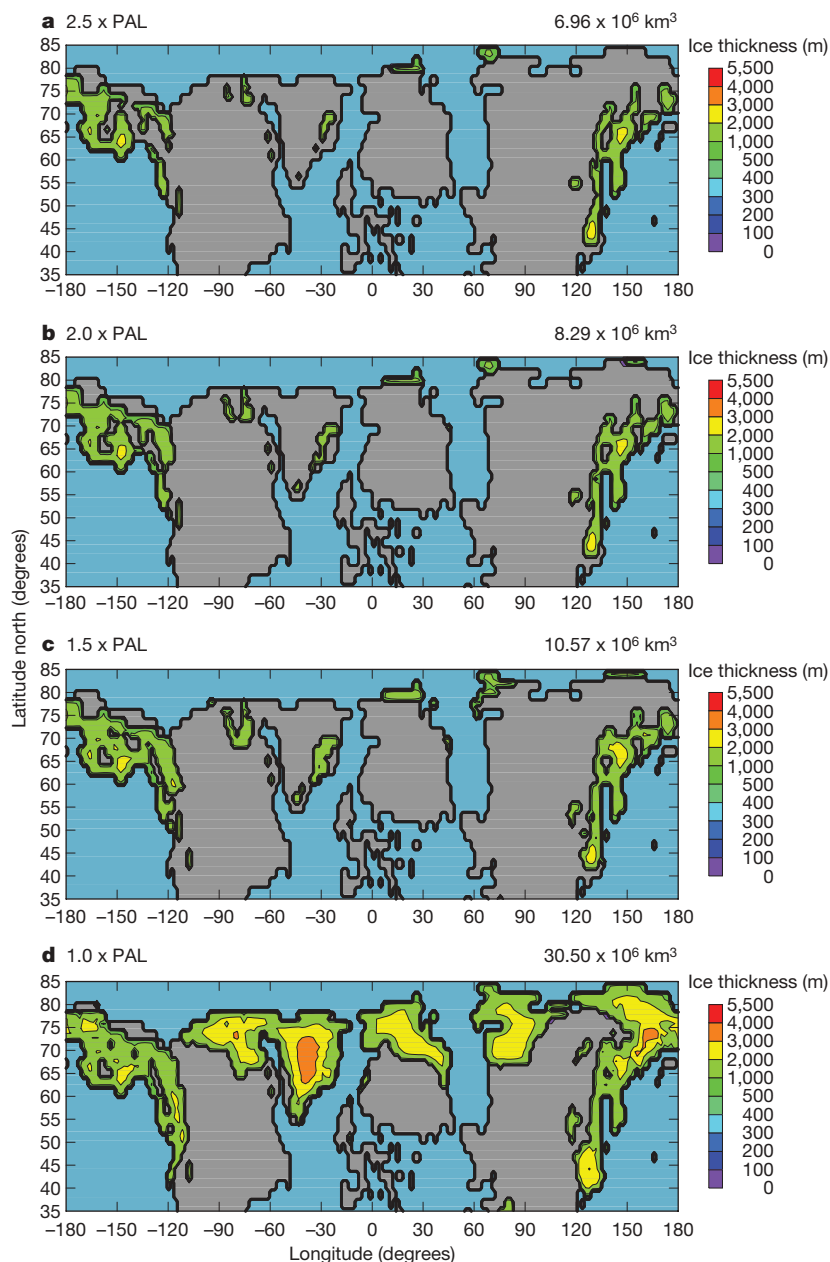
Simulated ice sheets, total ice volumes and associated changes in the mean isotopic composition of the ocean are shown in Figs 1 and 3, and in Supplementary Information. At relatively high levels of  $\text{CO}_2$  near the Antarctic glaciation threshold, small ice caps form on the highest elevations of western North America, northeast Asia, East Greenland and other locations where prevailing storm tracks intersect coastlines with steep relief. Despite the initial prescription of a favourable cool summer orbit, the onset of major glaciation does not occur until  $\text{CO}_2$  reaches pre-industrial levels (280 p.p.m.v.), with the largest continental ice sheet forming on Greenland because of its broad plateau and moist maritime climate. In simulations using lower topography (see Supplementary Information), major glaciation is delayed until  $\text{CO}_2$  drops below 180 p.p.m.v. Assuming that simulated northern-hemispheric ice sheets had an average isotopic composition similar to modern Greenland ice ( $-35\text{‰}$ )<sup>27</sup>, they would have enriched mean ocean  $\delta^{18}\text{O}_w$  by nearly  $0.8\text{‰}$ . This, in addition to Antarctica's model-derived contribution of  $0.5\text{‰}$  (Fig. 1), could produce most of the observed shift at Oi-1 without invoking deep-sea cooling; but this would only be possible if atmospheric  $\text{CO}_2$  fell to pre-industrial levels or below (Fig. 4).

In a situation of decreasing Cenozoic  $\text{CO}_2$ , our model first produces small isolated ice caps in the Antarctic interior during cold austral summer orbits at  $\text{CO}_2$  levels as high as  $6 \times \text{PAL}$  (ref. 22), but major Antarctic glaciation does not occur until  $\text{CO}_2$  reaches  $\sim 2.7 \times \text{PAL}$ . In the Northern Hemisphere, small ice caps appear on eastern Greenland and the highest elevations of the surrounding continents over a broad range of  $\text{CO}_2$  (Fig. 3), but major glaciation only occurs when  $\text{CO}_2$  falls near or below  $1 \times \text{PAL}$ . The lower  $\text{CO}_2$  threshold for the large Northern Hemisphere continents is due to their greater seasonality, lower latitudes and consequently warmer summers. Most proxy-based estimates for late Eocene to middle Oligocene  $\text{CO}_2$  range between 500 and 1,200 p.p.m.v. (ref. 10), well above our simulated Northern Hemisphere glaciation threshold (Fig. 4). Our results are consistent with the recent discovery of Eocene and Oligocene ice-rafted debris in the Greenland Sea<sup>8</sup> (Fig. 3a, b), but at these relatively high levels of  $\text{CO}_2$  they support a picture of small isolated ice caps and alpine outlet glaciers as the source rather than continental-scale ice sheets as recently suggested<sup>9</sup>.



**Figure 2 | Simulated isotopic composition of snowfall on a glaciating Antarctic continent.** **a**, Austral winter (June, July, August) isotopic averages (‰) on an ice-free Antarctica at  $3 \times \text{PAL}$ , just above the glaciation threshold. The austral winter average is shown for the  $3 \times \text{PAL}$  case because most summer precipitation in the ice-free case is rain. **b**, Annual mean isotopic

composition of precipitation at  $2.5 \times \text{PAL}$  and with an intermediate ice sheet (black outlines in continental interior). **c**, Same as **b** except with  $2 \times \text{PAL}$  and a fully developed early Oligocene East Antarctic ice sheet. Intermediate and fully glaciated ice-sheet geometries in **b** and **c** (extent and surface elevations) are taken from simulations in ref. 22.



**Figure 3 | Simulations of Northern Hemisphere ice sheets for progressively lower values of  $\text{CO}_2$ .** Simulations are for earliest Oligocene palaeogeography and favourable cold boreal summer orbit, and are made with a coupled GCM/ice-sheet model. Ice thicknesses are shown in metres, with

corresponding total Northern Hemisphere ice volumes ( $10^6 \text{ km}^3$ ) above the right corner of each panel. Palaeogeographical boundary conditions and an alternative set of simulations with continental elevations reduced by 50% are shown in the Supplementary Information.

For major bipolar glaciation to have occurred at Oi-1, CO<sub>2</sub> would first have to cross the Antarctic glaciation threshold ( $\sim 750$  p.p.m.v.) and then fall more than 400 p.p.m.v. within  $\sim 200$  kyr to reach the Northern Hemisphere threshold (Fig. 4). Increased sea ice and upwelling in the Southern Ocean<sup>13,29</sup> and falling sea level<sup>14</sup> could have acted as feedbacks accelerating CO<sub>2</sub> drawdown at the time of Oi-1. This is supported by CO<sub>2</sub> proxy records and carbon-cycle model results showing a drop in CO<sub>2</sub> across the Eocene/Oligocene transition<sup>10,13,14</sup>, but none of these reconstructions reach the low levels required for Northern Hemisphere glaciation. We therefore conclude that major bipolar glaciation at the Eocene/Oligocene transition is unlikely, and Mg/Ca-based estimates of deep-sea temperatures across the boundary<sup>5</sup> are unreliable. Our findings lend support to the hypothesis that the 1-km deepening of the carbonate compensation depth and the associated carbonate ion effect on deep-water calcite mask a cooling signal in the Mg/Ca records<sup>4,5</sup>. Therefore, the observed isotope shift at Oi-1 is best explained by Antarctic glaciation<sup>22</sup> accompanied by 4.0 °C of cooling in the deep sea or slightly less ( $\sim 3.3$  °C) if there was additional ice growth on West Antarctica (see Methods and Supplementary Information). This explanation is in better agreement with sequence stratigraphic estimates of sea-level fall at Oi-1 ( $70 \pm 20$  m)<sup>19,20</sup> equivalent to 70–120% of modern Antarctic ice volume, and coupled GCM/ice-sheet simulations showing 2–5 °C cooling and expanding sea ice in the Southern Ocean in response to Antarctic glaciation<sup>29</sup>. Additional support for ocean cooling is provided by new records from Tanzania<sup>16</sup> and the Gulf of Mexico<sup>15</sup>, where Mg/Ca temperature estimates show  $\sim 2.5$  °C cooling in shallow, continental shelf settings during the first step of the Eocene/Oligocene transition.

In summary, our model results show that the Northern Hemisphere contained glaciers and small, isolated ice caps in high elevations through much of the Cenozoic, especially during favourable orbital periods (Fig. 3a–c). However, major continental-scale Northern Hemisphere glaciation at or before the Oi-1 event (33.6 Myr) is unlikely, in keeping with recently published high-resolution Eocene

isotope records<sup>30</sup>. Proxy reconstructions of Cenozoic carbon dioxide<sup>10,11</sup> remain well above our model's threshold for Northern Hemisphere glaciation until around the Oligocene/Miocene boundary. Since that time, transient Northern Hemisphere ice sheets could have grown during favourable orbital periods and may help to account for the magnitude of Neogene isotope and sea-level variability<sup>17</sup> despite pronounced hysteresis in Antarctic ice-sheet dynamics<sup>18</sup>. The first major event to be considered in this context is Mi-1 ( $\sim 23.1$  Myr ago)<sup>3</sup>, an ephemeral 200-kyr isotopic excursion similar in magnitude to Oi-1 and coeval with a prolonged interval of low obliquity variance<sup>26</sup> favourable for ice-sheet development. Although no definitive evidence of widespread northern-hemispheric glaciation exists before  $\sim 2.7$  Myr ago, pre-Pliocene records from subsequently glaciated high northern latitudes are generally lacking. More highly resolved CO<sub>2</sub> records focusing on specific events, along with additional geological information from high northern latitudes, will help to unravel the Cenozoic evolution of the cryosphere. According to these results, this evolution may have included an episodic northern-hemispheric ice component for the past 23 million years.

## METHODS SUMMARY

The GCM and thermomechanical ice-sheet models are interactively coupled, whereby net annual surface mass balance on the ice sheet is calculated from monthly mean GCM meteorological fields of temperature and precipitation horizontally interpolated to the higher-resolution ice-sheet grid. Simulations in Fig. 3 were run to equilibrium (30 kyr) using a cold boreal summer orbit with high eccentricity (0.05), low obliquity (22.5°) and precession placing aphelion in July. The simulations producing large ice sheets (Fig. 3d, h) were repeated in asynchronous coupled mode<sup>22</sup> accounting for climate–ice feedbacks and time-continuous orbital forcing to confirm that the fixed-orbit results in Fig. 3 are representative of those with orbital variations.

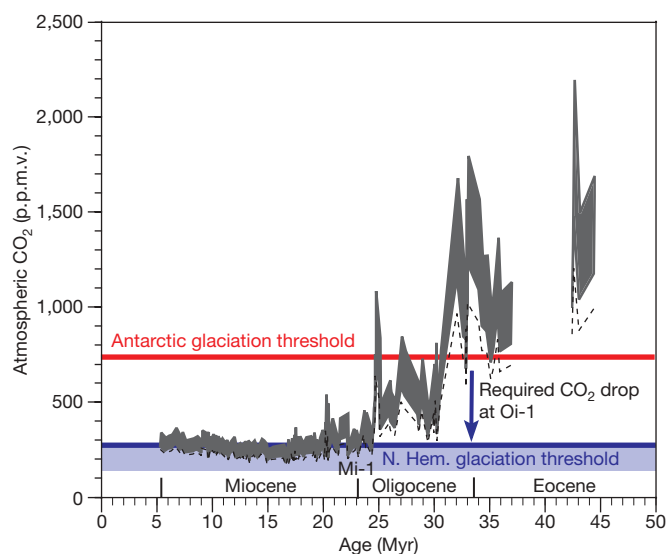
A modified version of the ice-sheet model accounting for floating ice shelves and migrating grounding lines was used to determine the potential for additional ice growth over West Antarctica at Oi-1. In this model version, the buttressing effect of an expanding proto-Ross ice shelf assists the growth of some additional ice, but only if ocean temperatures (and sub-ice melt rates) are assumed to be similar to modern (see Supplementary Fig. 1).

Ice volumes simulated by the ice-sheet model are converted to eustatic sea level according to the global ocean-area fraction in our 34 Myr palaeogeography (0.731). Equivalent  $\delta^{18}\text{O}_w$  (change in the average isotopic composition of the ocean) reflects either assumed isotopic ice compositions mentioned in the text, or those derived from the simulated isotopic composition of precipitation falling on the ice sheets using the stable isotope physics described previously<sup>28</sup>. In these calculations, the isotopic composition of the ocean was given a uniform global value of  $-1.2\text{‰}$ , consistent with ice-free conditions at the beginning of the experiment. With Antarctic ice at  $-35\text{‰}$ ,  $\delta^{18}\text{O}_w$  is 0.0246 per  $10^6 \text{ km}^3$  of grounded ice.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 3 April; accepted 12 August 2008.

- Barrett, P. J. Antarctic paleoenvironment through Cenozoic times: A review. *Terra Antart.* **3**, 103–119 (1996).
- Shackleton, N. J. *et al.* Oxygen isotope calibration of the onset of ice-raftering and history of glaciation in the North Atlantic region. *Nature* **307**, 620–623 (1984).
- Miller, K. G., Fairbanks, R. G. & Mountain, G. S. Tertiary oxygen isotope synthesis, sea level history, and continental margin erosion. *Paleoceanography* **1**, 1–20 (1987).
- Coxall, H. K., Wilson, P. A., Pälicke, H., Lear, C. & Backman, J. Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature* **433**, 53–57 (2005).
- Lear, C. H., Rosenthal, Y., Coxall, H. K. & Wilson, P. A. Late Eocene to early Miocene ice-sheet dynamics and the global carbon cycle. *Paleoceanography* **19**, PA4015, doi: 10.1029/2004PA001039 (2004).
- Billups, K. & Schrag, D. P. Application of benthic foraminiferal Mg/Ca ratios to questions of Cenozoic climate change. *Earth Planet. Sci. Lett.* **209**, 181–195 (2003).
- Zanazzi, A., Kohn, M. J., MacFadden, B. J. & Terry, D. O. Jr. Large temperature drop across the Eocene–Oligocene transition in central North America. *Nature* **445**, 639–642 (2007).
- Eldrett, J. S., Harding, I. C., Wilson, P. A., Butler, E. & Roberts, A. P. Continental ice in Greenland during the Eocene and Oligocene. *Nature* **466**, 176–179 (2007).
- Tripathi, A. *et al.* Evidence for Northern Hemisphere glaciation back to 44 Ma from ice-raftered debris in the Greenland Sea. *Earth Planet. Sci. Lett.* **265**, 112–122 (2008).



**Figure 4 | Model-generated CO<sub>2</sub> thresholds for Antarctic and Northern Hemisphere glaciation superposed on a Cenozoic record of atmospheric CO<sub>2</sub>.** The CO<sub>2</sub> record is taken from stable carbon isotopic values of diunsaturated alkenones<sup>10</sup>. The dashed line represents a lowermost limit, assuming  $\delta^{18}\text{O}$ -derived temperatures used in the calculation of CO<sub>2</sub> partial pressure are accurate, and the lower and upper bounds of the shaded (grey) area assume temperatures are 3 °C and 6 °C warmer, respectively. The blue arrow shows the drop in CO<sub>2</sub> required for Northern Hemisphere glaciation at Oi-1. The blue shading shows the range of uncertainty based on alternate Northern Hemisphere simulations with lower continental elevations (see text and Supplementary Information).



10. Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B. & Bohaty, S. M. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309**, 600–603 (2005).
11. Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide over the past 60 million years. *Nature* **406**, 695–699 (2000).
12. Laskar, J. *et al.* A long-term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004).
13. Zachos, J. & Kump, L. Carbon cycle feedbacks and the initiation of Antarctic glaciation in the earliest Oligocene. *Global Planet. Change* **47**, 51–66 (2005).
14. Merico, A., Tyrrell, T. & Wilson, P. A. Eocene/Oligocene ocean de-acidification linked to Antarctic glaciation by sea level fall. *Nature* **452**, 979–982 (2008).
15. Katz, M. E. *et al.* Stepwise transition from the Eocene greenhouse to the Oligocene icehouse. *Nature Geosci.* **1**, 329–334 (2008).
16. Lear, C., Bailey, T. R., Pearson, P. N., Coxhall, H. K. & Rosenthal, Y. Cooling and ice growth across the Eocene–Oligocene transition. *Geology* **36**, 251–354, doi:10.1130/G1124 (2008).
17. Pekar, S. & DeConto, R. M. High-resolution ice-volume estimates for the early Miocene: Evidence for a dynamic ice sheet in Antarctica. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **231**, 101–109 (2006).
18. Pollard, D. & DeConto, R. M. Hysteresis in Cenozoic Antarctic ice sheet variations. *Global Planet. Change* **45**, 9–21 (2005).
19. Pekar, S. F. & Christie-Blick, N. Resolving apparent conflicts between oceanographic and Antarctic climate records and evidence for a decrease in  $p\text{CO}_2$  during the Oligocene through early Miocene (34–16 Ma). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **260**, 41–49 (2008).
20. Kominz, M. A. & Pekar, S. F. Oligocene eustasy from two-dimensional sequence stratigraphic backstripping. *Geol. Soc. Am. Bull.* **113**, 291–304 (2001).
21. Kennett, J. P. Cenozoic evolution of Antarctic glaciation, the circum-Antarctic oceans and their impact on global paleoceanography. *J. Geophys. Res.* **82**, 3843–3859 (1977).
22. DeConto, R. M. & Pollard, D. Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric  $\text{CO}_2$ . *Nature* **421**, 245–249 (2003).
23. Larsen, H. C. *et al.* Seven million years of glaciation in Greenland. *Science* **264**, 952–955 (1994).
24. Raymo, M. E. & Ruddiman, W. F. Tectonic forcing of late Cenozoic climate. *Nature* **359**, 117–122 (1992).
25. St John, K. Cenozoic ice-rafting history of the central Arctic Ocean: terrigenous sands on the Lomonosov Ridge. *Paleoceanography* **23**, PA1505 (2008).
26. Zachos, J., Pagani, M., Sloan, L. & Thomas, E. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
27. Lhomme, N., Clarke, G. K. C. & Ritz, C. Global budget of water isotopes inferred from polar ice sheets. *Geophys. Res. Lett.* **32**, L20502, doi:10.1029/2005GL023774 (2005).
28. Mathieu, R. *et al.* Simulation of stable water isotope variations by the GENESIS GCM for modern conditions. *J. Geophys. Res.* **107**, doi:10.1029/2001JD900255 (2002).
29. DeConto, R. M., Pollard, D. & Harwood, D. Sea ice feedback and Cenozoic evolution of Antarctic climate and ice sheets. *Paleoceanography* **22**, PA3214, doi:10.1029/2006PA001350 (2007).
30. Edgar, K. M., Wilson, P. A., Sexton, P. F. & Suganuma, Y. No extreme bipolar glaciation during the Eocene calcite compensation shift. *Nature* **488**, 908–911 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This material is based on work supported by the National Science Foundation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.M.D. ([deconto@geo.umass.edu](mailto:deconto@geo.umass.edu)).

## METHODS

**Global climate/ice-sheet model.** The GCM and ice-sheet components of our model are the same as those used in prior simulations of Antarctic glaciation<sup>18,22</sup>, allowing interhemispheric comparisons of glaciation potential at different levels of atmospheric CO<sub>2</sub>. The horizontal resolution of the atmospheric component of the GCM is T31 (~3.75° by ~3.75°) with 18 vertical layers. Surface models including a 50-m slab ocean, dynamic–thermodynamic sea ice, and multi-layer models of snow, soil and vegetation are on a finer 2° × 2° grid. The GCM is coupled to a thermomechanical ice-sheet model. The ice model grid over Antarctica is polar stereographic with a resolution of 40 km by 40 km, and 1° latitude by 1° longitude over the Northern Hemisphere continents. Ice-sheet evolution is driven by surface mass balance forcing from the GCM. Terrestrial ice flow is modelled using the shallow ice approximation, while accounting for internal ice temperatures, basal sliding, bedrock isostatic relaxation and lithospheric flexure.

Monthly mean meteorological fields (temperature and precipitation) used in the calculation of net-annual surface mass balance are horizontally interpolated from the GCM to the higher-resolution ice-sheet grids, using a lapse-rate adjustment to account for local topographic offsets between model components. A positive degree–day parameterization with an imposed diurnal cycle is used to calculate ablation while accounting for refreezing of meltwater. Mass balance is re-calculated every 200 model years to account for evolving surface elevations.

All simulations in Fig. 3 were run to equilibrium using a cold boreal summer orbit with high eccentricity (0.05) and low obliquity (22.5°), with the longitude of precession placing aphelion in July. Simulations producing large ice sheets were repeated in asynchronous coupled mode, with the GCM rerun every 10,000 ice-model years to account for changing albedo, topography and evolving orbital parameters (see Supplementary Information Fig. 4). This was done to confirm that summer warming during unfavourable phases of the precession cycle was insufficient to stop the onset of glaciation initiated during a cold boreal summer orbit, and that the results shown in Fig. 3 are representative of those with orbital variations. Although the CO<sub>2</sub>–glaciation thresholds shown here are model-dependent, the GCM's sensitivity to 2 × PAL (2.5 °C) is average among models used in future and palaeoclimate modelling studies, and the use of a different atmospheric component with similar CO<sub>2</sub> sensitivity is expected to produce similar results.

The Oi-1 Antarctic glaciation experiment<sup>22</sup> was repeated using a modified version of the ice-sheet model with extensions accounting for floating ice shelves and migrating grounding lines<sup>31</sup> to determine the potential for additional ice growth over West Antarctica. In this model, a combined set of scaled equations for sheet and shelf flow accounts for both horizontal shear ( $\partial u/\partial z$ ) and stretching

( $\partial u/\partial x$ ) dominant in grounded and floating ice, respectively. In this simulation, the buttressing effect of an expanding proto-Ross ice shelf aids the growth of additional West Antarctic ice. If ocean temperatures and sub-ice melt rates are assumed to be similar to modern, our model builds only an additional  $7.5 \times 10^6 \text{ km}^3$  of ice on West Antarctica (Supplementary Information Fig. 1), not enough to ameliorate the Oi-1 amplitude problem significantly. Further expansion of grounding lines to the continental shelf as occurred at the Last Glacial Maximum<sup>32</sup> is unlikely, owing to warmer conditions in the early Oligocene. Equilibrium ice-free bedrock elevations for the model are obtained from modern observed bathymetry<sup>33</sup>, isostatically rebounded with modern ice removed. If West Antarctic bedrock elevations were higher in the early Oligocene<sup>34</sup> with more land area above sea level and shallower continental shelves, this could have allowed somewhat greater amounts of early West Antarctic ice than in Supplementary Information Fig. 1, which in turn would require less ocean cooling to explain the magnitude of Oi-1.

**Water isotopes.** The water isotope tracer model<sup>28</sup> passively tracks the hydrological cycle in the GCM atmosphere and applies relevant fractionation physics during phase transitions. The model considers <sup>1</sup>H<sub>2</sub><sup>18</sup>O and <sup>1</sup>HD<sup>16</sup>O and accounts for evaporative, condensational and post-condensational processes. Reservoir effects are differentiated over ocean, land (vegetation) and ice sheets. In the case of our Palaeogene Antarctic glaciation experiments, the isotopic composition of the ocean was given a uniform global value of −1.2‰. In modern control simulations<sup>28</sup>, the seasonal and spatial distribution of δ<sup>18</sup>O of precipitation generated by the model is close to observed values, except in the highest elevations of the Antarctic interior, where, as in most GCMs, model surface temperatures are warmer than observed and the δ<sup>18</sup>O of precipitation is 5–10‰ too heavy. This bias is not relevant for the initially ice-free conditions and smaller ice sheets in the Palaeogene simulations, because of their lower topography and warmer temperatures than modern.

31. Pollard, D. & DeConto, R. M. in *Glacial Sedimentary Processes and Products* (eds Hambrey, M. et al.), 37–52 (Internat. Assoc. Sedimentologists Spec. Publ. 39, Blackwell, 2007).
32. Huybrechts, P. Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. *Quat. Sci. Rev.* **21**, 203–231 (2002).
33. Bamber, J. A. & Bindshadler, R. A. An improved elevation dataset for climate and ice-sheet modelling: validation with satellite imagery. *Ann. Glaciol.* **25**, 439–444 (1997).
34. Sorlien, C. C. et al. Oligocene development of the West Antarctic ice sheet recorded in eastern Ross Sea strata. *Geology* **35**, 467–470 (2007).

# Crystallographic preferred orientation of akimotoite and seismic anisotropy of Tonga slab

Rei Shiraishi<sup>1</sup>, Eiji Ohtani<sup>1</sup>, Kyuichi Kanagawa<sup>3</sup>, Akira Shimojuku<sup>1,4</sup> & Dapeng Zhao<sup>2</sup>

The mineral akimotoite, ilmenite-structured  $\text{MgSiO}_3$ , exists at the bottom of the Earth's mantle transition zone and within the uppermost lower mantle, especially under low-temperature conditions<sup>1</sup>. Akimotoite is thought to be a major constituent of the harzburgite layer of subducting slabs, and the most anisotropic mineral in the mantle transition zone<sup>2–4</sup>. It has been predicted that if akimotoite crystals are preferentially oriented by plastic deformation, a cold subducted slab would be extremely anisotropic<sup>5</sup>. However, there have been no studies of crystallographic preferred orientations and very few reports of plastic deformation experiments for  $\text{MgSiO}_3$  ilmenite. Here we present plastic deformation experiments on polycrystalline akimotoite, which were conducted at confining pressures of 20–22 GPa and temperatures of 1,000–1,300 °C. We found a change in crystallographic preferred orientation pattern of akimotoite with temperature, where the *c*-axis maximum parallel to the compression direction develops at high temperature, whereas the *c* axes are preferentially oriented parallel to the shear direction or perpendicular to the compression direction at lower temperature. The previously reported difference in compressional-wave seismic anisotropy between the northern and southern segments of the Tonga slab at depths of the mantle transition zone<sup>6</sup> can conceivably be attributed to the difference in the crystallographic preferred orientation pattern of akimotoite at varying temperature within the slab.

A polycrystalline akimotoite used for the present deformation experiments was synthesized at high pressure and temperature with a Kawai-type multi-anvil apparatus at Tohoku University. The synthesized polycrystalline akimotoite was confirmed to have no crystallographic preferred orientation (CPO).

The akimotoite specimen was then deformed by either uniaxial compression or simple shear geometry at high pressures and temperatures with the Kawai-type multi-anvil apparatus. The experimental conditions and results are given in Table 1.

We performed two types of experiment (types I and II; Table 1). In both type I and type II experiments, pressure was increased at room temperature, and temperature was then increased at the desired

pressure. In type I experiments the sample was annealed for either 10 or 60 min, and then quenched. In type II experiments the sample was further deformed by slightly increasing the pressure, and then quenched. Two blank experiments were also conducted: a cold-compression experiment (DI01) and a non-annealing experiment (DI08). In run DI01, pressure was increased and then decompressed immediately. In run DI08, temperature was increased quickly (over about 10 min) to 1,200 °C after increasing the pressure, and the sample was subsequently quenched without annealing.

Recovered samples were cut in half parallel to the compression direction. Thin sections were then prepared and were polished with a colloidal silica suspension. A thin (~10 nm) coating of carbon was applied to decrease specimen charging. In each sample, the crystallographic orientations of 166–271 akimotoite grains were determined by the electron backscatter diffraction (EBSD) technique<sup>7</sup> (Table 1).

The change in sample thickness revealed the total compressional strains of samples deformed by uniaxial compression. The shear strain of specimen DI07 was calculated from the axial displacement of the pistons. The total compressional strains were 0.1–0.3. DI07 was deformed to a shear strain of 0.6. The microstructures and grain sizes (~10 µm) of samples, except specimen DI07, were similar. The grain size of DI07 was about 5 µm, slightly smaller than the other samples.

Equal-area lower-hemisphere projections of akimotoite  $\langle 1120 \rangle$ ,  $\langle 10\bar{1}0 \rangle$  and  $[0001]$  directions in the samples, deformed at 1,000–1,300 °C, are given in Fig. 1, where the compression and shear directions are shown by black arrows. Deformed polycrystalline  $\text{MgSiO}_3$  akimotoite at 1,200–1,300 °C (Fig. 1a–c) has a CPO characterized by a strong *c*-axis maximum subparallel to the compression direction, and  $\langle 1120 \rangle$  and  $\langle 10\bar{1}0 \rangle$  axis girdles normal to the compression direction. This CPO suggests a dominant slip system with glide on (0001), which is in good agreement with the observation<sup>8</sup> that the dominant slip system in experimentally deformed akimotoite is  $1/3 \langle 1120 \rangle$  (0001). In addition, the basal glide on (0001) was reported in some analogue of akimotoite, such as the trigonal structure of ilmenite<sup>9</sup>. In specimens deformed at 1,000 °C (Fig. 1d, e), the *c* axes

**Table 1 | Summary of experimental conditions and results**

Run no.	Pressure (GPa)	<i>P</i> – <i>T</i> path*	Assembly	Temperature (°C)	Heating duration (min)	Compressional strain	Shear strain	Grain size (µm)	CPO
DI01	21.1	–	Compression	–	–	0.16	–	7.4	Random
DI02	21.1	Type I	Compression	1,200	60	0.09	–	8.0	(0001) $\perp \sigma_1$
DI03	20.0→20.5	Type II	Compression	1,300	140	0.09	–	7.7	(0001) $\perp \sigma_1$
DI04	21.1	Type I	Compression	1,200	10	0.14	–	10.0	(0001) $\perp \sigma_1$
DI06	22.2→22.5	Type II	Compression	1,000	160	0.29	–	8.2	[0001] $\perp \sigma_1$
DI07	22.2	Type I	Shear	1,000	60	–	0.61	5.8	[0001]    $\sigma_1$
DI08	21.1	–	Compression	1,200	0	0.14	–	8.0	Random

\* We examined two types of pressure–temperature (*P*–*T*) path. In type I experiments, samples were annealed under the target conditions and then quenched. In type II experiments, the sample was further deformed by increasing the pressure slightly.  $\perp \sigma_1$ , perpendicular to the compression direction; ||  $\sigma_1$ , parallel to the shear direction.

<sup>1</sup>Institute of Mineralogy, Petrology, and Economic Geology. <sup>2</sup>Department of Geophysics, Tohoku University, Sendai 980-8578, Japan. <sup>3</sup>Department of Earth Sciences, Chiba University, Chiba 263-8522, Japan. <sup>4</sup>Department of Earth and Planetary Sciences, Faculty of Sciences, Kyushu University, Fukuoka 812-8581, Japan.



are preferentially orientated parallel to the shear direction or perpendicular to the compression direction. The CPO of the sample deformed in uniaxial compression is not axially symmetric, suggesting a deviation of its deformation from uniaxial geometry. The texture of the sample deformed in simple shear (DI07) is axially symmetric about the shear direction. Both  $\langle 11\bar{2}0 \rangle$  and  $\langle 10\bar{1}0 \rangle$  axes spread along a girdle around the  $c$ -axis maximum. This CPO therefore suggests the dominance of slip in the  $[0001]$  direction on multiple planes. In contrast, no clear CPO was developed in the sample deformed at room temperature (DI01) or in the non-annealing sample (DI08). From these blank tests, the CPOs observed in samples DI02–DI07 are considered to have developed during the plastic deformation at the target pressures and temperatures.

In addition, there is no difference in CPO pattern resulting from the difference between the  $P$ – $T$  paths in the type I and type II experiments, judging from the fact that the CPO pattern of DI02 is the same as that of DI03.

Thus, observed akimotoite CPOs and inferred dominant slip systems differ in their deformation temperatures. Slip in the  $[0001]$  direction is probably dominant at a lower temperature (1,000 °C), whereas the basal glide on (0001) becomes dominant at higher temperatures (1,200–1,300 °C). The fabric transition occurs at about 1,100 °C. Fabric transitions with increasing temperature are also reported in other minerals such as quartz<sup>10–12</sup> and olivine<sup>13–15</sup>. For wet quartz it has been reported<sup>16</sup> that there is a possible transition

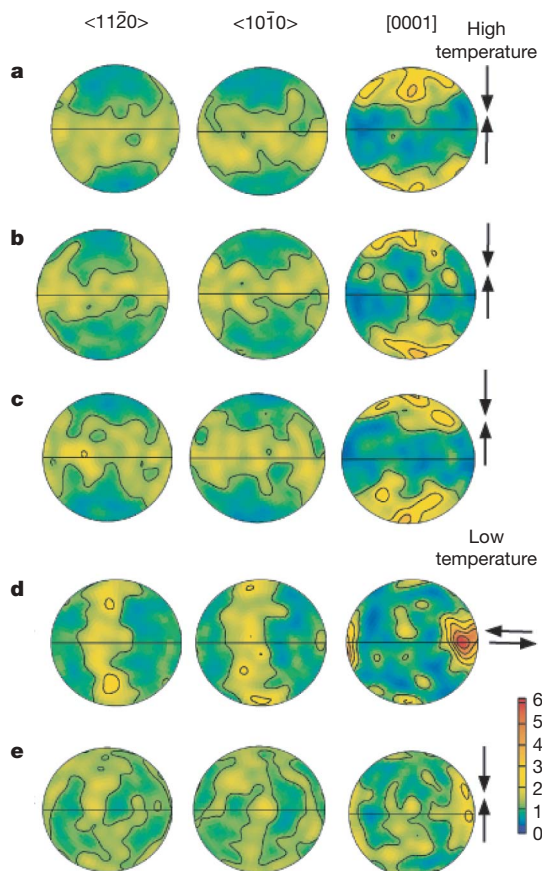
from  $\langle 11\bar{2}0 \rangle$  (0001) to  $[0001]\{11\bar{2}0\}$  with increasing temperature. This observation is similar to that of akimotoite observed in this study.

We calculated seismic wave velocities from the akimotoite CPO data to examine the relationship between akimotoite CPO and seismic anisotropy. We used the elastic constants and density of akimotoite under the mantle transition zone conditions determined previously<sup>4</sup> with the molecular dynamic approach. The Voigt–Reuss–Hill average was used to calculate the seismic anisotropy. We used the program Anis2k (ref. 17) to calculate bulk elastic constants  $C_{ij}$  from the CPO data, as well as  $P$ -wave velocities. For the CPO data of DI03 and DI06 deformed in uniaxial compression, we randomly rotated the orientation data about the compression axis five times and used all rotated data for the following calculation, to decrease the deviations of those CPO data from uniaxial symmetry. Single-crystal akimotoite has a  $V_P$  anisotropy of 14.4%, which is shown in Fig. 2a. The results calculated from the CPO data for three representative samples (DI03, DI06 and DI07) are shown in Fig. 2b–d.

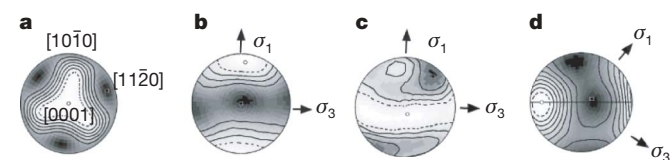
For the sample deformed at a higher temperature (DI03), the  $V_P$  anisotropy is 3.0%. In the sample deformed at a lower temperature, the  $V_P$  anisotropy of the compression experiment (DI06) and the simple shear experiment (DI07) are 1.0% and 4.3%, respectively. As regards other mantle transition-zone minerals, it has been reported<sup>18</sup> that the  $V_P$  anisotropies of 60% wadsleyite and 40% garnet deformed to shear strains of 1.0 and 0.5 are 2% and 1%, respectively. Although there are no CPO data for the other mantle transition-zone minerals, the anisotropy of a rock composed of 100% akimotoite is at least fourfold to fivefold that of a rock composed of 60% wadsleyite and 40% garnet. Akimotoite therefore has a much greater effect on the seismic anisotropy of subducting slabs at transition-zone depths.

Because of the difference in CPO pattern between the sample deformed at 1,300 °C (DI03) and that deformed at 1,000 °C (DI06 and DI07), the anisotropy pattern also depends on temperature. The  $P$  wave propagates most slowly in the shear direction or in the direction perpendicular to the compression direction at 1,000 °C, but in the compression direction at 1,300 °C. This is because the velocity of the  $P$  wave through an akimotoite single crystal is slowest in the  $c$ -axis direction (Fig. 2a).

The spatial variation of seismic anisotropy in the Tonga subducting slab was shown recently<sup>6</sup>. The slab is divided into two segments: the northern segment at latitudes 17–19° S and the southern segment at latitudes 19.5–27° S (Fig. 3). The magnitude of the anisotropy is 5–7% for  $P$  waves and 9–12% for  $S$  waves, and the direction of maximum velocity is different in each of the two slab segments. In the northern segment,  $P$  waves propagate more slowly in the slab normal direction. In contrast,  $P$  waves propagate more slowly in



**Figure 1 | Equal-area projections of pole figures for  $\langle 11\bar{2}0 \rangle$ ,  $\langle 10\bar{1}0 \rangle$  and  $[0001]$  directions of akimotoite in all samples. a, DI02 ( $T = 1,200$  °C;  $n = 271$ ); b, DI03 ( $T = 1,300$  °C;  $n = 197$ ); c, DI04 ( $T = 1,200$  °C;  $n = 220$ ); d, DI07 ( $T = 1,000$  °C;  $n = 166$ ); e, DI06 ( $T = 1,000$  °C;  $n = 216$ );  $n$  is the number of grains measured. The projections are coloured according to the density of data points and are contoured at multiples of uniform distribution as shown in the scale at the bottom right. The north–south direction corresponds to the compression direction in a–c and e, and to the shear-plane normal direction in d. Pairs of bold arrows represent the compression direction or the shear direction.**



**Figure 2 | P-wave anisotropies calculated using Anis2k.  $V_P$  contours are shown; black squares, maximum velocities ( $V_{\max}$ ); white circles, minimum velocities ( $V_{\min}$ ). a, Akimotoite single crystal.  $V_{\max} = 12.44$  km s<sup>−1</sup>;  $V_{\min} = 10.77$  km s<sup>−1</sup>. b, Akimotoite aggregate experimentally deformed at a relatively high temperature (1,300 °C) by uniaxial compression (DI03).  $V_{\max} = 11.65$  km s<sup>−1</sup>;  $V_{\min} = 11.31$  km s<sup>−1</sup>. c, d, Experimental deformation at a relatively low temperature (1,000 °C) by uniaxial compression (c; DI06;  $V_{\max} = 11.59$  km s<sup>−1</sup>;  $V_{\min} = 11.47$  km s<sup>−1</sup>) and by simple shear (d; DI07;  $V_{\max} = 11.67$  km s<sup>−1</sup>;  $V_{\min} = 11.18$  km s<sup>−1</sup>). The north–south direction corresponds to the compression direction in b and c, and to the shear-plane normal direction in d. The horizontal line and the east–west direction in d correspond to the shear plane and the shear direction, respectively. The  $\sigma_1$  direction for the deformed samples is the direction of the advancing pistons.**

the slab sinking direction in the southern segment. As regards S waves, the directions of maximum and minimum velocities do not coincide with the slab normal and sinking directions. We compared P-wave anisotropy in experimentally deformed akimotoite aggregates with that observed in the Tonga subducting slab as mentioned above (Figs 2 and 3). We assumed that the maximum compressive direction ( $\sigma_1$ ) and the minimum compressive direction ( $\sigma_3$ ) in experimentally deformed samples corresponded to the orientation of the principal stress axes estimated from focal mechanisms of deep earthquakes in the Tonga slab<sup>6</sup>. In the southern Tonga slab segment, the P-wave velocity is slower and faster in the  $\sigma_1$  and  $\sigma_3$  directions, respectively (Fig. 3c), which is similar to that in the akimotoite aggregate deformed at 1,300 °C (Fig. 2b). In contrast, in the northern Tonga slab segment, the P-wave velocity is faster in the  $\sigma_2$  direction and slowest in the direction of the bisector between the  $\sigma_1$  and  $\sigma_3$  directions (Fig. 3b), which correlates well with that in the akimotoite aggregate deformed at 1,000 °C (Fig. 2d). Thus, the difference in seismic anisotropy between the northern and southern Tonga slab segments is attributable to the difference in CPO patterns of akimotoite resulting from differences in temperature. However, the transition temperature in the Tonga slab between the southern and northern segments could not be determined quantitatively because CPO patterns are also dependent on strain rate, and the geological strain rates are much smaller than the experimental strain rates. The geometry of the Tonga slab is complicated, specifically at greater depths<sup>19,20</sup> (Fig. 3). In addition, it has been reported that there is a difference in the distribution of the low-velocity zones in the mantle wedge between the northern and southern parts of the Tonga back-arc<sup>21</sup>. The low-velocity zone in the deep mantle wedge above the southern part of the Tonga slab may be caused by partial melting or by the existence of fluids from dehydration of the slab. These observations suggest that the temperature of the southern part of the Tonga slab is higher than that of the northern part. There are probably lateral variations in temperature in the Tonga slab that give

rise to the difference in the seismic anisotropy pattern in the Tonga slab. It has been suggested that there are strong lateral variations in  $V_P$  and  $V_S$ , and that these variations are caused by a petrological anomaly, such as compositional or mineralogical variation<sup>22</sup>. In addition to compositional variation, the change in CPO pattern with temperature may have contributed to the differences between the northern and southern segments of the Tonga slab. Because ringwoodite and majorite, which may also be major constituents in the lower part of the mantle transition zone, are almost isotropic, the CPO of akimotoite must control the seismic anisotropy of the slab in the transition zone.

## METHODS SUMMARY

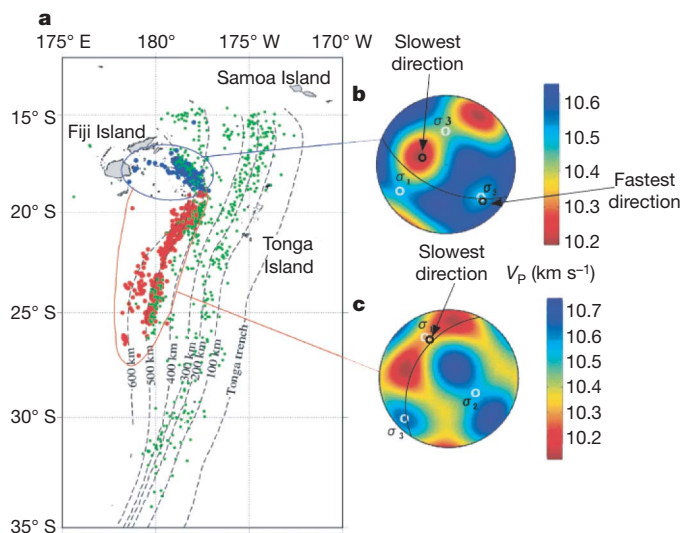
**Experimental procedure.** The furnace assembly was composed of a sintered  $\text{ZrO}_2$  pressure medium (an octahedron with an edge length of 10 mm), Ta electrodes and a  $\text{LaCrO}_3$  heater. Temperature was measured with W3%Re–W25%Re thermocouples. The starting material, which was placed in a platinum capsule, was an  $\text{MgSiO}_3$  glass fabricated from oxides. Synthesis experiments were conducted at 20–22 GPa and 1,250–1,550 °C for 60 min.

We measured the water content of the starting material by Fourier-transform infrared spectroscopy with a JASCO MFT-2000 instrument. The water content was determined by integrating the infrared absorption spectrum from 3,200 to 3,750  $\text{cm}^{-1}$  using a previous calibration of the extinction coefficient<sup>23</sup>. The water content in akimotoite is 24 p.p.m. by weight, which is extremely low compared with hydrous akimotoite<sup>24</sup>.

The sample was sandwiched between hard alumina pistons inserted in the furnace assembly to produce differential stresses during compression. To minimize the deformation during cold compression, crushable alumina was placed at the outer ends of the pistons. Crushable alumina is initially very porous and soft. However, it becomes dense and works as good piston material on compression. These ideas and the cell assemblies were based on ref. 25.

**EBSD measurement.** EBSD patterns were obtained using a Nordlys II EBSD detector mounted on a Jeol JSM-6460 scanning electron microscope at Chiba University, operating with an accelerating voltage of 20 kV and a beam current of 1.5–2.4 nA, and indexed manually with Channel 5 software (Flamenco) from HKL Technology.

Received 23 October 2007; accepted 1 August 2008.



**Figure 3 | P-wave anisotropy of the Tonga slab and the deformed akimotoite.** **a**, Map showing the geometry of the subducting Tonga slab and epicentres of earthquakes. **b**, **c**, P-wave anisotropies of the northern (**b**) and southern (**c**) Tonga slab segments<sup>6</sup>. Green dots in **a** represent earthquake foci at depths between 100 and 500 km; blue and red dots show earthquakes at depths greater than 500 km in the northern and southern Tonga segments. The dashed lines are equal-depth contours of the Tonga slab<sup>26</sup>. The  $V_P$  contour diagrams (**b**, **c**) are equal-area lower-hemisphere projections, in which the white circles show the directions of the stress axes ( $\sigma_1$ ,  $\sigma_2$  and  $\sigma_3$ ) in the Tonga slab deduced from the focal mechanism solutions, and the black arc lines show the intersection of the Tonga slab with the hemisphere at depths greater than 500 km. Black circles show the fastest and the slowest  $V_P$  directions expected from the deformed akimotoite.

1. Akaogi, M., Tanaka, A. & Ito, E. Garnet-ilmenite-perovskite transitions in the system  $\text{Mg}_4\text{Si}_4\text{O}_{12}$ – $\text{Mg}_3\text{Al}_2\text{Si}_3\text{O}_{12}$  at high-pressures and high-temperatures: phase equilibria, calorimetry and implications for mantle structure. *Phys. Earth Planet. Inter.* **132**, 303–324 (2002).
2. Weidner, D. J. & Ito, E. Elasticity of  $\text{MgSiO}_3$  in the ilmenite phase. *Phys. Earth Planet. Inter.* **40**, 65–70 (1985).
3. Da Silva, C. R. S., Karki, B. B., Stixrude, L. & Wentzcovitch, R. M. *Ab initio* study of the elastic behavior of  $\text{MgSiO}_3$  ilmenite at high-pressure. *Geophys. Res. Lett.* **26**, 943–946 (1999).
4. Zhang, Y., Zhao, D. & Matsui, M. Anisotropy of akimotoite: A molecular dynamics study. *Phys. Earth Planet. Inter.* **151**, 309–319 (2005).
5. Anderson, D. L. *Theory of the Earth* (Blackwell, 1989).
6. Vavryčuk, V. Spatially dependent seismic anisotropy in the Tonga subduction zone: A possible contributor to the complexity of deep earthquakes. *Phys. Earth Planet. Inter.* **155**, 63–72 (2006).
7. Randle, V. *Microtexture Determination and its Applications* 2nd edn (Maney, 2003).
8. Cordier, P. in *Plastic Deformation of Minerals and Rocks* (eds Karato, S. I. & Wenk, H. R.) 137–179 (American Mineralogical Society, 2002).
9. Bascou, J., Raposo, M. I. B., Vauchez, A. & Egydio-Silva, M. Titanohematite lattice-preferred orientation and magnetic anisotropy in high-temperature mylonites. *Earth Planet. Sci. Lett.* **198**, 77–92 (2002).
10. Lister, G. S. Fabric transitions in plastically deformed quartzites: competition between basal, prism and rhomb systems. *Bull. Mineral.* **102**, 232–241 (1979).
11. Schmid, S. M. & Casey, M. Complete fabric analysis of some commonly observed quartz c-axis patterns. *Am. Geophys. Un. Geophys. Monogr.* **36**, 263–286 (1986).
12. Mainprice, D., Bouchez, J.-L., Blumenfeld, P. & Tubia, J. M. Dominant c slip in naturally deformed quartz; implications for dramatic plastic softening at high temperature. *Geology* **14**, 2181–2202 (1986).
13. Katayama, I. & Karato, S. Effect of temperature on the B- to C-type olivine fabric transition and implication for flow pattern in subduction zones. *Phys. Earth Planet. Inter.* **157**, 33–45 (2006).
14. Carter, N. L. & Ave'Lallemant, H. G. High temperature flow of dunite and peridotite. *Geol. Soc. Am. Bull.* **81**, 2181–2202 (1970).
15. Jung, H. & Karato, S.-I. Water-induced fabric transitions in olivine. *Science* **293**, 1460–1463 (2001).
16. Blacic, J. D. Plastic deformation mechanisms in quartz: The effect of water. *Tectonophysics* **27**, 271–294 (1975).
17. Mainprice, D. A. FORTRAN program to calculate seismic anisotropy from the lattice preferred orientation of minerals. *Comput. Geosci.* **16**, 385–393 (1990).

18. Tommasi, A., Mainprice, D., Cordier, P., Thoraval, C. & Couvy, H. Strain-induced seismic anisotropy of wadsleyite polycrystals and flow patterns in the mantle transition zone. *J. Geophys. Res.* **109**, B12406, doi:10.1029/2005JB004168 (2004).
19. Chen, W.-P. & Brudzinski, M. R. Evidence for a large-scale remnant of subducted lithosphere beneath Fiji. *Science* **292**, 2475–2479 (2001).
20. Chen, W.-P. & Brudzinski, M. R. Seismic anisotropy in the mantle transition zone beneath Fiji–Tonga. *Geophys. Res. Lett.* **30**, 1682, doi:10.1029/2002GL016330 (2003).
21. Zhao, D. *et al.* Depth extent of the Lau back-arc spreading center and its relation to subduction processes. *Science* **278**, 254–257 (1997).
22. Brudzinski, M. R. & Chen, W.-P. A petrologic anomaly accompanying outboard earthquakes beneath Fiji–Tonga: Corresponding evidence from broadband P and S waveforms. *J. Geophys. Res.* **108**, B62299, doi:10.1029/2002JB002012 (2003).
23. Paterson, M. S. The determination of hydroxyl by infrared absorption in quartz, silicate glasses and similar materials. *Bull. Mineral.* **105**, 20–29 (1982).
24. Bolfan-Casanova, N., Keppler, H. & Rubie, D. C. Water partitioning between nominally anhydrous minerals in the MgO–SiO<sub>2</sub>–H<sub>2</sub>O system up to 24 GPa: implications for the distribution of water in the Earth's mantle. *Earth Planet. Sci. Lett.* **182**, 209–221 (2000).
25. Karato, S. & Rubie, D. C. Toward an experimental study of deep mantle rheology: A new multianvil sample assembly for deformation studies under high pressures and temperatures. *J. Geophys. Res.* **102**, 20111–20122 (1997).
26. Gudmundsson, O. & Sambridge, M. A regionalized upper mantle (RUM) seismic model. *J. Geophys. Res.* **103**, 7121–7136 (1998).

**Acknowledgements** This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Science, Sport, and Technology of the Japanese Government.

**Author Contributions** R.S. performed experiments and took the lead in writing the manuscript. E.O. and A.S. designed the study. K.K. performed EBSD analyses. D.Z. worked on the seismological aspects of this study. All co-authors took part in the discussion and interpretation of the results and improving the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.S. ([siraisir@ganko.tohoku.ac.jp](mailto:siraisir@ganko.tohoku.ac.jp)).

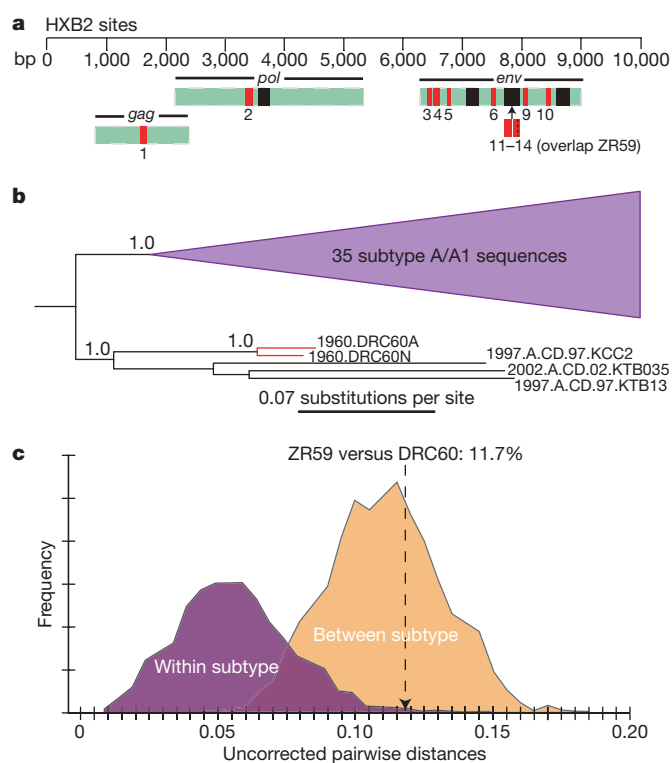


# Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960

Michael Worobey<sup>1</sup>, Marlea Gemmel<sup>1</sup>, Dirk E. Teuwen<sup>2,3</sup>, Tamara Haselkorn<sup>1</sup>, Kevin Kunstman<sup>4</sup>, Michael Bunce<sup>5</sup>, Jean-Jacques Muyembe<sup>6,7</sup>, Jean-Marie M. Kabongo<sup>6</sup>, Raphaël M. Kalengayi<sup>6</sup>, Eric Van Marck<sup>8</sup>, M. Thomas P. Gilbert<sup>1†</sup> & Steven M. Wolinsky<sup>4</sup>

Human immunodeficiency virus type 1 (HIV-1) sequences that pre-date the recognition of AIDS are critical to defining the time of origin and the timescale of virus evolution<sup>1,2</sup>. A viral sequence from 1959 (ZR59) is the oldest known HIV-1 infection<sup>1</sup>. Other historically documented sequences, important calibration points to convert evolutionary distance into time, are lacking, however; ZR59 is the only one sampled before 1976. Here we report the amplification and characterization of viral sequences from a Bouin's-fixed paraffin-embedded lymph node biopsy specimen obtained in 1960 from an adult female in Léopoldville, Belgian Congo (now Kinshasa, Democratic Republic of the Congo (DRC)), and we use them to conduct the first comparative evolutionary genetic study of early pre-AIDS epidemic HIV-1 group M viruses. Phylogenetic analyses position this viral sequence (DRC60) closest to the ancestral node of subtype A (excluding A2). Relaxed molecular clock analyses incorporating DRC60 and ZR59 date the most recent common ancestor of the M group to near the beginning of the twentieth century. The sizeable genetic distance between DRC60 and ZR59 directly demonstrates that diversification of HIV-1 in west-central Africa occurred long before the recognized AIDS pandemic. The recovery of viral gene sequences from decades-old paraffin-embedded tissues opens the door to a detailed palaeovirological investigation of the evolutionary history of HIV-1 that is not accessible by other methods.

We screened 27 tissue blocks (8 lymph node, 9 liver and 10 placenta) obtained from Kinshasa between 1958 and 1960 by polymerase chain reaction with reverse transcription (RT-PCR); one lymph node biopsy specimen contained HIV-1 RNA. Viral nucleic acids were extracted from this specimen using protocols optimized for the recovery of nucleic acids from ancient or degraded samples<sup>3,4</sup>. After reverse transcription, 12 out of the 14 short HIV-1 complementary DNA fragments in the study (Fig. 1a) were amplified by PCR using a panel of conserved primer pairs from different regions of the viral genome (Supplementary Table 1). Each PCR product was cloned and sequenced. Sequences were reproducible after repeated extractions and were not the result of PCR contamination (see Fig. 1a and Supplementary Table 1 for fragment designations). The results were confirmed independently in two laboratories (Fig. 1b and Supplementary Fig. 1), with the second laboratory successfully identifying the positive 1960 specimen in a blinded assay. The short fragments of the 1960 sample were found to be of subtype A and not to be a mosaic of contemporary sequences (see Supplementary Information for a detailed discussion of the authenticity of the 1960



**Figure 1 | Fragments amplified from DRC60, and the results of the phylogenetic and sequence analyses.** **a**, The HIV-1 genome fragments that were successfully amplified from DRC60 (red) and are available for ZR59 (black). The numbering for the HIV-1 sequences corresponds to the HXB2 reference sequence (Supplementary Table 1). **b**, The A/A1 subtree from the unconstrained (in which a molecular clock is not enforced) BMCMC phylogenetic analysis. Supplementary Fig. 1 depicts the complete phylogenetic tree (50% majority rule consensus tree of the posterior sample, with branch lengths averaged across the sample). Posterior probabilities are shown on nodes with support >0.95. 1960.DRC60A is the University of Arizona consensus sequence, and 1960.DRC60N is the Northwestern University consensus sequence (that is, the sequences independently recovered in each of the two laboratories). The DRC60 sequences form a strongly supported clade with three modern sequences also sampled in the DRC. **c**, Smoothed histograms of within-subtype (A2, A/A1, B, C, D, F1, F2, H, J, K) and between-subtype distances.

<sup>1</sup>Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA. <sup>2</sup>Sanofi Pasteur, F-69367 Lyon Cedex 07, France. <sup>3</sup>UCB SA Pharma, Braine l'Alleud, BE-1420, Belgium. <sup>4</sup>The Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA. <sup>5</sup>Ancient DNA Laboratory, School of Biological Sciences and Biotechnology, Murdoch University, Perth, Western Australia 6150, Australia. <sup>6</sup>Department of Anatomy and Pathology, University of Kinshasa, Kinshasa B.P. 864, Democratic Republic of the Congo. <sup>7</sup>National Institute for Biomedical Research, National Laboratory of Public Health, Kinshasa B.P. 1197, Democratic Republic of the Congo. <sup>8</sup>Department of Pathology, University Hospital, University of Antwerp, Antwerp B-2610, Belgium. †Present address: Centre for Ancient Genetics, Biological Institute, University of Copenhagen, Copenhagen DK-2100, Denmark.

sequences). Consensus nucleotide sequences from these short HIV-1 fragments were concatenated for study. The analyses included reference sequences from the Los Alamos National Laboratory HIV sequence database and sequences recovered as part of this study from three paraffin-embedded tissue specimens collected from AIDS patients in Belgium and Canada between 1981 and 1997.

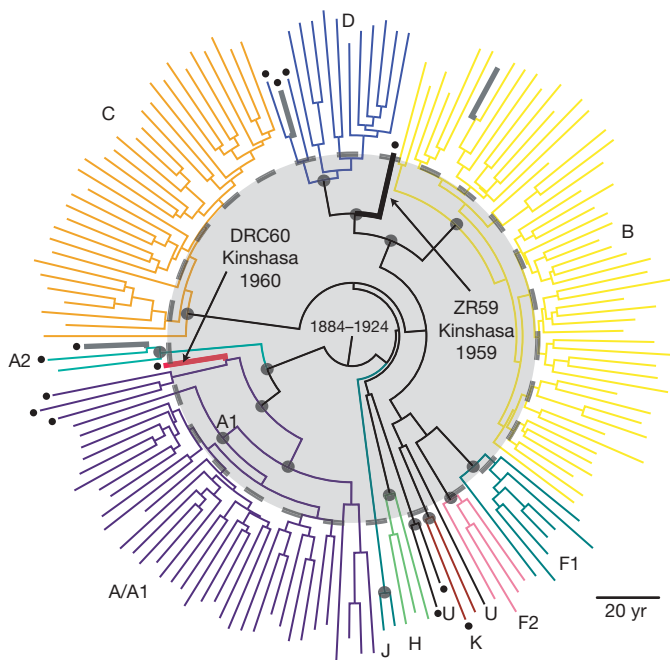
HIV-1 sequences were analysed in MrBayes v3.1.2 (ref. 5) using an unconstrained (in which a molecular clock was not enforced) Bayesian Markov chain Monte Carlo (BMCMC) method. The phylogenetic analyses confirmed that the DRC60 consensus sequences from the two laboratories were derived from a single patient (uncorrected pairwise distance of 1.4%). The sequences were positioned close to the ancestral node of the subtype A lineage (excluding sub-subtype A2), forming a monophyletic clade with three modern sequences from the DRC (Fig. 1b and Supplementary Fig. 1). Assuming a similar rate of evolution along all branches on a tree, the divergence between two sequences reflects the time elapsed since their shared ancestor. As predicted, the DRC60 sequences had a shorter branch length to the A/A1 ancestral node than the contemporary subtype A viruses sampled from the same geographic region ( $P = 1.0$ ).

We validated the time of origin of the 1960 sequence by comparisons of the predicted date to the documented date. With the DRC60 date treated as an unknown, we calculated an evolutionary rate on the basis of the distribution of branch lengths on the unconstrained phylogenetic trees sampled by MrBayes. To limit the effects of evolutionary rate differences between clades and uncertainties in rooting the HIV-1 M group phylogeny, we focused on the subtype A/A1 subtree (Supplementary Fig. 1) and analysed root-to-tip branch lengths relative to the sampling year. The mean estimates for the year of origin of the DRC60 consensus sequences from the University of Arizona and Northwestern University laboratories were 1959 (95% highest probability distribution (HPD) 1902–1984) and 1959 (95% HPD 1915–1985), respectively, corroborating the authenticity of the DRC60 sequences and the existence of a clock-like signal in our data set (see later). Despite initial indications that recombination might seriously confound phylogenetic dating estimates<sup>6</sup>, subsequent work has suggested that recombination is not likely to systematically bias HIV-1 dates in one direction or the other, although it is expected to increase variance<sup>7</sup>. The close match between the predicted and the actual dates of both ZR59 (ref. 2) and DRC60 provides support for this view and gives an unambiguous indication that HIV-1 evolves in a fairly reliable clock-like fashion.

The uncorrected pairwise distance between DRC60 and ZR59 in their overlapping *env* region was 11.7% (Fig. 1c). This genetic distance is greater than 99.2% of within-subtype comparisons (within-subtype difference, range 0.01–0.15; between-subtype difference, range 0.05–0.18). Because each subtype represents several decades of independent evolution in the human population<sup>2,8</sup>, the extensive divergence between DRC60 and ZR59 indicates that the HIV-1 M group founder virus began to diversify in the human population (and that HIV-1 probably entered Kinshasa) decades before 1960.

We applied a relaxed clock BMCMC coalescent framework as implemented in BEAST v1.4.7 (ref. 9) to estimate the time to the most recent common ancestor (TMRCA) of the HIV-1 M group. This approach robustly incorporates phylogenetic uncertainty and accounts for the possibility of variable substitution rates among lineages and differences in the demographic history of the virus, sampling phylogenies and parameter estimates in proportion to their posterior probability<sup>10</sup>. As with other studies of HIV-1 (ref. 11), comparisons of the marginal likelihoods of strict versus relaxed clock models (both of which are implemented in BEAST) indicated overwhelming support for relaxed clocks (data available on request). Hence, the use of strict clock models with these data would be inappropriate and would probably yield misleadingly small error estimates with regard to both timing and substitution rates.

Using substitution rates calibrated with sequences sampled at different time points, we obtained a posterior distribution of rooted tree topologies with branch lengths in unit time (Fig. 2 and Supplementary Fig. 2). The median estimated substitution rate for the concatenated subregions of the *gag-pol-env* genes was  $2.47 \times 10^{-3}$  substitutions per site per year (95% HPD  $1.90\text{--}2.95 \times 10^{-3}$ ). The inclusion of the 1959 and 1960 sequences seemed to improve estimation of the TMRCA of the M group (Table 1), limiting the influence of the coalescent tree prior on the posterior TMRCA distributions compared with the data set that excluded these earliest cases of HIV-1. With DRC60 and ZR59 included, the different demographic/coalescent models gave highly consistent results, with tighter and more similar date ranges compared with the analyses that excluded them and 95% HPDs that extend no later than 1933. The best-fit model incorporated a constant population size demographic model (TMRCA 1921, 95% HPD 1908–1933). The model with a general, non-parametric prior (the Bayesian skyline plot tree prior)<sup>12,13</sup> that indicated a more complex (and biologically plausible) demographic history (Supplementary Fig. 3) had a statistically indistinguishable degree of support (TMRCA 1908, 95% HPD 1884–1924). Moreover, the population expansion demographic model<sup>9</sup>, which was a slightly worse fit to the data compared with the constant population and Bayesian skyline plot models, could not be rejected given the Bayes factor comparison of models (Table 1). The inability to strongly reject the model with a constant population size prior is counterintuitive because it is clear that the HIV-1 population size has increased notably. We speculate that this finding might be due to the simplest model providing a good fit to a relatively short,



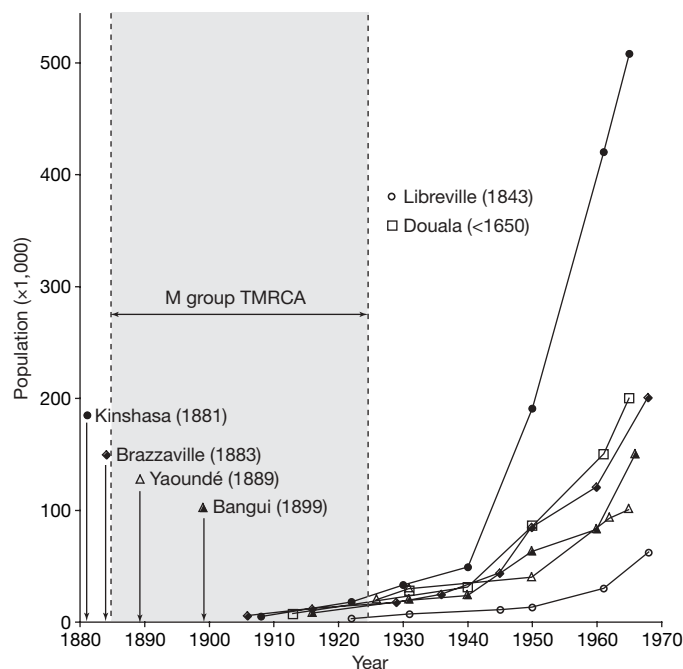
**Figure 2 | Maximum clade credibility topology inferred using BEAST v1.4.7 under a Bayesian skyline plot tree prior.** Branch lengths are depicted in unit time (years) and represent the median of those nodes that were present in at least 50% of the sampled trees. DRC60 (red), ZR59 (black) and the three control sequences from paraffin-embedded specimens from known AIDS patients (grey) are depicted in bold. The 95% HPD of the TMRCA is indicated at the root of the tree. Nodes (sub-subtype and deeper) with posterior probability of 1.0 are marked with grey circles. Unclassifiable strains are labelled 'U'. Sequences sampled in the DRC are highlighted with a bullet at the tip. DRC60 and the two control sequences from the DRC each form monophyletic clades with previously published sequences from the DRC, whereas the Canadian control sequence clusters, as expected, with subtype B sequences. The dashed circle and shaded area show the extensive HIV-1 diversity in Kinshasa in the 1950s. Supplementary Fig. 2 shows the tree in rectangular form with taxon labels.

information-poor alignment, in comparison with more parameterized models.

Acid-containing fixatives such as Bouin's solution can cause base modifications of nucleic acids, leading to the generation of erroneous bases in sequences derived from such samples<sup>3</sup>. However, the replication of all sequences from independent PCR amplifications and the uncorrected pairwise distance between the consensus sequences from the two laboratories (1.4%) suggest that few of the mutations on the DRC60 lineage are damaged-induced. Moreover, our relaxed clock methods are likely to be fairly robust to the presence of such mutations in one lineage<sup>9</sup>. Nevertheless, additional old sequence data would be helpful for resolving what impact, if any, this possible source of error had on the slightly earlier dates we calculated compared with previous estimates that did not include early calibration points<sup>2,8,14,15</sup>. Interestingly, the best-fit model for the data set that excluded ZR59 and DRC60 (Table 1) gave a TMRCA estimate of 1933 (1919–1945), which is very similar to that of ref. 2. This suggests that the inclusion of the old sequences, rather than the vagaries associated with a much shorter alignment than that analysed by ref. 2, might explain the discrepancy. Also, one earlier study, using sequences from the DRC only<sup>16</sup>, produced dating and demography estimates very similar to ours. Overall, there is broad agreement between all of these studies in spite of differences in data and methods.

Our estimation of divergence times, with an evolutionary timescale spanning several decades, together with the extensive genetic distance between DRC60 and ZR59 indicate that these viruses evolved from a common ancestor circulating in the African population near the beginning of the twentieth century; TMRCA dates later than the 1930s are strongly rejected by our statistical analyses. The topology of the HIV-1 group M phylogeny provides further support for this conclusion. Unlike ZR59, which is basal to subtype D<sup>1</sup>, DRC60 branches off from the ancestral node of subtype A/A1 (Fig. 2 and Supplementary Figs 1 and 2). Thus, it is clear that phylogenetically distinct subtypes (and/or their progenitors) were already present in the DRC by this early time point (Fig. 2). Notably, DRC60 and ZR59 cluster with other strains from the same geographical region and basal to other members of their respective subtypes, a pattern consistent with the hypothesis that the subtypes spread through lineage founder effects worldwide, whereas a more diverse array of forms remained at the site of origin in Africa<sup>17,18</sup>.

The reservoir of the ancestral virus still exists among wild chimpanzee communities in the same area on the African continent<sup>19</sup>. Humans acquired a common ancestor of the HIV-1 M group by cross-species transmission under natural circumstances<sup>20</sup>, probably predation<sup>21</sup>. The Bayesian skyline plot (Supplementary Fig. 2), which tracks effective population size through time, suggests that HIV-1 group M experienced an extensive period of relatively slow growth in the first half of the twentieth century. A similar pattern has been inferred using sequences sampled only in the DRC<sup>16</sup>. This pattern, and the short duration between the first presence of urban agglomerations in this area and the timing of the most recent common ancestor of HIV-1 group M (Fig. 3), suggests that the rise of cities may have facilitated the initial establishment and the early spread of HIV-1. Hence, the founding and growth of colonial administrative and trading centres such as Kinshasa<sup>22</sup> may have enabled the region to become the epicentre of the HIV/AIDS pandemic<sup>23</sup>.



**Figure 3 | The origin and growth of the major settlements near the epicentre of the HIV-1 group M epidemic.** In the countries surrounding the putative zone of cross-species transmission<sup>19</sup> (current-day Cameroon, Central African Republic, DRC, Republic of Congo, Gabon and Equatorial Guinea) there was not a single site with a population exceeding 10,000 until after 1910. The founding date of each major city in the region is listed beside its name. Most were founded only shortly before the estimated TMRCA of group M. The demographic data are from ref. 23.

The archival banks of Bouin's-fixed paraffin-embedded tissue specimens accumulated by many hospitals in west-central Africa provide a vast source of clinical material for viral genetic analysis. As with the 1918 Spanish influenza pandemic virus<sup>24,25</sup>, a deep perspective on the evolutionary history of HIV-1 using sequences resurrected from the earliest cases in Africa could yield important insights into the pathogenesis, virulence and evolution of pandemic AIDS viruses.

## METHODS SUMMARY

A total of 813 Bouin's-fixed paraffin-embedded histopathological blocks were recovered from the 1958–1962 archives of the Department of Anatomy and Pathology at the University of Kinshasa. The boxes were stored until transfer to the University of Arizona, where 8 lymph node, 9 liver and 10 placenta samples from 1958–1960 were selected for RNA preservation analysis and HIV-1 RNA screening. We used a human  $\beta$ -2-microglobulin (*B2M*) quantitative RT-PCR assay to assess RNA quality as described<sup>3</sup>. Digestion and extraction of these samples, and of three modern positive-control samples, were performed using QIAamp DNA micro kits (Qiagen) using the protocol described in ref. 3. We used 14 primer sets designed to anneal to highly conserved regions of the *gag*, *pol* and *env* genes of HIV-1 group M and to amplify very short fragments likely to be present even in ancient and/or degraded specimens (Supplementary Table 1). Reverse transcription was performed using the SuperScript III System for RT-PCR (Invitrogen). The cDNA was amplified by PCR using Platinum Taq HiFi enzyme (Invitrogen) and cloned using the TOPO TA Cloning Kit (Invitrogen). We constructed an alignment including 156 published reference sequences plus

**Table 1 | HIV-1 M group TMRCA estimates from BEAST analyses under different coalescent tree priors**

Coalescent tree prior	DRC60 and ZR59 excluded*	DRC60 and ZR59 included
Constant	<b>1933 (1919–1945)†, 0.0</b>	<b>1921 (1908–1933)†, 0.0</b>
Exponential	1907 (1874–1932), $-3.5 \pm 0.8$	1914 (1891–1930), $-2.1 \pm 1.5$
Expansion	1882 (1834–1917), $-2.7 \pm 0.8$	<b>1902 (1873–1922)†, <math>-1.6 \pm 1.5</math></b>
Logistic	1913 (1880–1937), $-2.3 \pm 0.8$	1913 (1891–1930), $-3.2 \pm 1.5$
Bayesian skyline plot	1882 (1831–1916), $-2.7 \pm 0.8$	<b>1908 (1884–1924)†, <math>-0.4 \pm 1.5</math></b>

Shown for each coalescent tree prior is the median, with the 95% highest probability distribution of TMRCA in parentheses. Also shown is the  $\log_{10}$  Bayes factor difference in estimated marginal likelihood ( $\pm$  estimated standard error) compared with the coalescent model with strongest support.

\*Concatenated *gag-pol-env* fragments available for either or both of ZR59 and DRC60 (994 nucleotides total, 507 from DRC60).

†TMRCA for the best-fit model and models not significantly worse than it are written in bold.



the sequences recovered in this study, concatenating the 12 (out of 14) fragments successfully amplified from the 1960 sample and the 4 fragments already available from the 1959 sample (994 bases total). We performed an unconstrained (not enforced by a molecular clock) BMCMC analysis in MrBayes v3.1.2 (ref. 5) and used the resulting MCMC sample to test whether the 1960 sequence exhibited properties consistent with its provenance (both age and geography). We used a relaxed molecular clock model, as implemented in BEAST v1.4.7 (ref. 9), to estimate the TMRCA of HIV-1 group M using the 1960 and 1959 samples and to investigate the demographic history of the virus. We also performed pairwise comparisons within and between subtypes for the 163 bases available for both DRC60 and ZR59.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 21 May; accepted 8 September 2008.**

- Zhu, T. F. *et al.* An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**, 594–597 (1998).
- Korber, B. *et al.* Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**, 1789–1796 (2000).
- Gilbert, M. T. P. *et al.* The isolation of nucleic acids from fixed, paraffin-embedded tissues — which methods are useful when? *PLoS ONE* **2**, e537 (2007).
- Worobey, M. Phylogenetic evidence against evolutionary stasis and natural abiotic reservoirs of influenza A virus. *J. Virol.* **82**, 3769–3774 (2008).
- Huelsenbeck, J. P. & Ronquist, F. MrBayes: Bayesian inference of phylogeny. *Bioinformatics* **17**, 754–755 (2001).
- Worobey, M. A novel approach to detecting and measuring recombination: insights into evolution in viruses, bacteria, and mitochondria. *Mol. Biol. Evol.* **18**, 1425–1434 (2001).
- Lemey, P. *et al.* The molecular population genetics of HIV-1 group O. *Genetics* **167**, 1059–1068 (2004).
- Gilbert, M. T. P. *et al.* The emergence of HIV-1 in the Americas and beyond. *Proc. Natl Acad. Sci. USA* **104**, 18566–18570 (2007).
- Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
- Salemi, M., de Oliveira, T., Ciccozzi, M., Rezza, G. & Goodenow, M. M. High-resolution molecular epidemiology and evolutionary history of HIV-1 subtypes in Albania. *PLoS ONE* **3**, e1390 (2008).
- Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol. Biol. Evol.* **18**, 1001–1013 (2001).
- Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
- Sharp, P. M. *et al.* The origins of acquired immune deficiency syndrome viruses: where and when? *Phil. Trans. R. Soc. Lond. B* **356**, 867–876 (2001).
- Salemi, M. *et al.* Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution. *FASEB J.* **15**, 276–278 (2001).
- Yusim, K. *et al.* Using human immunodeficiency virus type 1 sequences to infer historical features of the acquired immune deficiency syndrome epidemic and human immunodeficiency virus evolution. *Phil. Trans. R. Soc. Lond. B* **356**, 855–866 (2001).
- Vidal, N. *et al.* Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *J. Virol.* **74**, 10498–10507 (2000).
- Rambaut, A., Robertson, D. L., Pybus, O. G., Peeters, M. & Holmes, E. C. Human immunodeficiency virus phylogeny and the origin of HIV-1. *Nature* **410**, 1047–1048 (2001).
- Keele, B. F. *et al.* Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**, 523–526 (2006).
- Worobey, M. in *Global HIV/AIDS Medicine* (eds Volberding, P. A., Sande, M. A., Lange, J. & Greene, W. C.) 13–21 (Saunders Elsevier, 2008).
- Hahn, B. H., Shaw, G. M., De Cock, K. M. & Sharp, P. M. AIDS as a zoonosis: scientific and public health implications. *Science* **287**, 607–614 (2000).
- Hance, W. A. *Population, Migration, and Urbanization in Africa* 209–297 (Columbia Univ. Press, 1970).
- Chitnis, A., Rawls, D. & Moore, J. Origin of HIV type 1 in colonial French Equatorial Africa? *AIDS Res. Hum. Retrov.* **16**, 5–8 (2000).
- Taubenberger, J. K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889–893 (2005).
- Tumpey, T. M. *et al.* Characterization of the reconstructed 1918 Spanish Influenza pandemic virus. *Science* **310**, 77–80 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Wertheim and M. Sanderson for computational assistance, and L. Jewel for providing the Canadian control specimen. The NIH/NIAID and the David and Lucile Packard Foundation funded the research.

**Author Contributions** M.W., D.E.T., S.M.W. and M.T.P.G. designed the study. M.G., T.H., K.K. and M.T.P.G. performed digestion and extraction, PCR, quantitative PCR, cloning and sequencing experiments. M.T.P.G., M.G. and M.B. optimized DNA/RNA isolation methods and designed PCR assays. D.E.T., J.-J.M., E.V.M., J.-M.M.K. and R.M.K. organized and provided samples. M.W. analysed the data, performed the phylogenetic analyses, and wrote the paper. S.M.W. contributed to the analyses and writing. All authors discussed the results and commented on the manuscript.

**Author Information** The sequences reported in this study have been deposited in GenBank under accession numbers EU580739–EU580854 and EU589211–EU589236. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.W. ([worobey@email.arizona.edu](mailto:worobey@email.arizona.edu)).

## METHODS

**Archival samples.** Each individual block carried an original paper identification number permanently embedded in the paraffin. Laboratory books listed the corresponding identification numbers sequentially, and included the patient's age, sex, department of hospitalization, tissue type and date of sampling. Block identification number, sampling date and tissue type were transcribed onto an Excel spreadsheet, and the blocks were indexed, transferred into plastic boxes and photographed.

The results of the quantitative RT-PCR assay indicated that the integrity of the RNA preserved in these 27 samples ranged from moderate to undetectable, a range typical of Bouin's-fixed specimens<sup>3</sup>. The human RNA found in the 1960 lymph node biopsy sample that was found to be HIV-1-RNA-positive was of relatively good quality. The  $C_t$  values (quantitative PCR data available from the authors on request) were as low or lower (better) than more recent (1980–1990) paraffin-embedded tissues that have yielded short HIV-1 RNA amplicons<sup>3</sup>.

Three formalin-fixed paraffin-embedded necropsy specimens were obtained: a Canadian patient who died in 1997 (CAN97); a Congolese woman who died in Belgium in 1981 and who was retrospectively identified as an AIDS patient (BE81); and a Congolese man who died in Belgium in 1985 (BE85). The latter two cases were presumably infected in Zaire (now the Democratic Republic of the Congo). The phylogenetic reconstruction shows that their viral sequences are most closely related to modern sequences from the Democratic Republic of the Congo, whereas the Canadian specimen yielded a subtype B sequence, as predicted (Figs 1 and 2 and Supplementary Figs 1 and 2).

**RNA isolation and reverse transcription.** Between 5 and 10 microtome sections, 5–10  $\mu\text{m}$  in thickness, or an approximately equivalent amount of tissue shaved from each block with a disposable scalpel blade, were used for each digestion and extraction, as described<sup>3</sup>. Rigorous attention was given to preventing cross-contamination between samples by cleaning the outer surface of each block with a bleach solution, using fresh microtome/scalpel blades for each sectioning of each block, discarding the first few (exposed-surface) sections, and by performing the work in a room physically isolated from any human or HIV-1 PCR-product DNA. A 48-h digestion period (24 h at 65 °C, 24 h at 75 °C) was used. Post extraction nucleic acids were eluted into 100  $\mu\text{l}$  elution buffer AE and stored frozen at –80 °C until required for analyses.

Reverse transcription was performed simultaneously for the *gag*, *pol* and human B2M RNA fragments; *env* fragments 3–10; and *env* fragments 11–14 (Supplementary Table 1). This was performed with SuperScript III used according to the manufacturer's instructions. The protocol was as described<sup>3</sup> except that alternating 50 °C and 55 °C incubation periods of 30 min were used for a total of 6 h.

**Amplification, cloning and DNA sequencing.** The cDNA was PCR amplified in 25- $\mu\text{l}$  reactions, using 0.1  $\mu\text{l}$  Platinum Taq HiFi enzyme (Invitrogen), 250  $\mu\text{M}$  dNTP mix, 2 mM  $\text{MgSO}_4$ , 1 $\times$  PCR buffer, 0.4  $\mu\text{M}$  per primer, and 2  $\mu\text{l}$  cDNA for the *gag* and *pol* reactions or 1  $\mu\text{l}$  for the *env* ones, with annealing temperatures of 60 °C (*gag*, 60 cycles) or 55 °C (*pol*, 50 cycles; *env*, 55 cycles). Full details are available from the authors on request.

After amplification, the PCR-product DNA was visualized by agarose gel electrophoresis and then purified using Zymoclean DNA Clean and Concentrator-25 spin tubes (Zymo Research Corporation). PCR-product DNA was inserted into vector pCR2.1-TOPO using the TOPO TA Cloning Kit (Invitrogen). The University of Arizona Genomic Analysis and Technology Core Facility resolved the DNA sequence of the vector inserts on an Applied Biosystems 3730xl DNA analyser using ABI Big Dye 3.1 chemistry (Applied Biosystems). Nearly identical protocols were followed for the independent replication of the DRC60 results at Northwestern University.

**Alignments.** We downloaded the 2006 full-length HIV-1 sequence alignment from the Los Alamos National Laboratories HIV sequence database<sup>26</sup>. We retained only non-recombinant HIV-1 group M A–K subtype sequences (excluding G) and removed sequences suspected a priori of unusual evolutionary dynamics (such as those associated with the intravenous drug user epidemic in Eastern Europe and those with *nef* deletions, both of which exhibit abnormally slow evolutionary rates). We also reduced the size of the subtype B and C clades, which are heavily over-sampled relative to the others, by keeping only the first 5 sequences from any year/country pair and then randomly removing sequences until the sample size was similar to that of the other subtypes. This procedure left a total of 156 sequences. We then manually aligned the consensus sequence from

the 12 regions amplified from DRC60, plus the 4 regions available for ZR59, to the full-length sequences. These short regions (Fig. 1a and Supplementary Table 1) were then concatenated into an alignment 994 nucleotides in length. The four *env* fragments from DRC60 that overlapped with available data from ZR59 were concatenated into an alignment 163 nucleotides in length. Matching alignments with DRC60 and ZR59 removed were also constructed. All the alignments are available from the authors on request.

**MrBayes analyses.** We used a general time-reversible nucleotide substitution model with gamma-distributed rate heterogeneity among sites and performed four independent runs of 20 million steps, sampling every 2,000 steps. Examination of the MCMC samples with Tracer v1.4 (ref. 9) indicated convergence and adequate mixing of the Markov chain with estimated sample sizes in the thousands. We discarded the first 2 million steps from each run as burn-in, and combined the resulting MCMC samples for subsequent estimation of posteriors. The 50% majority rule consensus tree (Supplementary Fig. 1) is shown rooted on the branch identified by the rooted-tree method in BEAST v1.4.7 (ref. 9), described below; however, the group M rooting was not relevant to any dating analysis. We also estimated phylogenies using the same data set under neighbour-joining and maximum likelihood methods and the same substitution model. The DRC60 sequences fell in the same topological position as with the BMCMC methods, with short root-to-tip genetic distances, consistent with the MrBayes results (Supplementary Fig. 1). All data and trees available from the authors on request.

We used the posterior tree sample to test the hypothesis that the terminal nodes of the DRC60 sequences were closer to the inferred A/A1 ancestral node by calculating the proportion of sampled trees where the A/A1 node-to-tip distances were smaller for these sequences than for the three modern sequences from the DRC in the same clade (Fig. 1b).

To predict the date of sampling on the basis of the phylogenetic properties of the DRC60 sequences, we also plotted the branch lengths (A/A1 node to tips) against the time of sampling for all A/A1 sequences excluding DRC60 and calculated the best fit for the linear regression of genetic divergence against the year of sampling of the viruses<sup>27</sup>. We calculated the mean and 95% HPD of the predicted sampling date of each DRC60 consensus sequence on the basis of its node-to-tip distance and the inferred regression line calculated for each of 100 trees sampled by MrBayes.

**Bayesian MCMC inference of phylogeny using BEAST v1.4.7.** We used the Bayesian methods described previously<sup>9,10</sup>, which allow for the co-estimation of phylogeny and divergence times under a 'relaxed' molecular clock model, as implemented in BEAST v1.4.7 (ref. 9). All analyses were performed under an uncorrelated lognormal relaxed molecular clock model, using a general time-reversible nucleotide substitution model with heterogeneity among sites modelled with a gamma distribution. We investigated each demographic model (constant population, exponential growth, expansion growth, logistic growth) as well as a Bayesian skyline plot coalescent tree prior<sup>13</sup>, a general, non-parametric prior that enforces no particular demographic history. We used a piecewise linear skyline model with 10 groups. We then compared the marginal likelihoods for each model using Bayes factors estimated in Tracer v1.4 as described<sup>12,15</sup>. Bayes factors represent the ratio of the marginal likelihoods of the models being compared. A large ratio can indicate that one model is a significantly better fit to the data than another. We assessed the strength of the evidence that the best-fit model was superior to the others as described<sup>15</sup>.

For each analysis, two independent runs of 50 million steps were performed. Examination of the MCMC samples with Tracer v1.4 indicated convergence and adequate mixing of the Markov chains, with estimated sample sizes in the hundreds or thousands. After inspection with Tracer, we discarded an appropriate number of steps from each run as burn-in, and combined the resulting MCMC tree samples for subsequent estimation of posteriors. We summarized the MCMC samples using the maximum clade credibility topology found with TreeAnnotator v1.4.7 (ref. 9), with branch length depicted in years (median of those branches that were present in at least 50% of the sampled trees; Fig. 2). The Bayesian skyline plot was reconstructed using the posterior tree sample and Tracer v1.4.

26. Leitner, T. *et al.* HIV Sequence Compendium (<http://www.hiv.lanl.gov>) (Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, 2005).

27. Drummond, A., Pybus, O. G. & Rambaut, A. Inference of viral evolutionary rates from molecular sequences. *Adv. Parasitol.* **54**, 331–358 (2003).

# Individual differences in non-verbal number acuity correlate with maths achievement

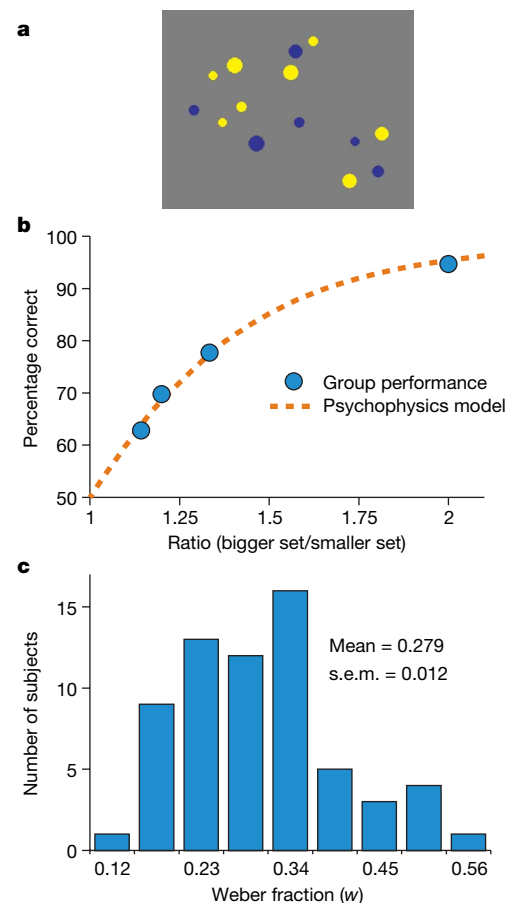
Justin Halberda<sup>1</sup>, Michèle M. M. Mazzocco<sup>1,2</sup> & Lisa Feigenson<sup>1</sup>

Human mathematical competence emerges from two representational systems. Competence in some domains of mathematics, such as calculus, relies on symbolic representations that are unique to humans who have undergone explicit teaching<sup>1,2</sup>. More basic numerical intuitions are supported by an evolutionarily ancient approximate number system that is shared by adults<sup>3–6</sup>, infants<sup>7</sup> and non-human animals<sup>8–13</sup>—these groups can all represent the approximate number of items in visual or auditory arrays without verbally counting, and use this capacity to guide everyday behaviour such as foraging. Despite the widespread nature of the approximate number system both across species and across development, it is not known whether some individuals have a more precise non-verbal ‘number sense’ than others. Furthermore, the extent to which this system interfaces with the formal, symbolic maths abilities that humans acquire by explicit instruction remains unknown. Here we show that there are large individual differences in the non-verbal approximation abilities of 14-year-old children, and that these individual differences in the present correlate with children’s past scores on standardized maths achievement tests, extending all the way back to kindergarten. Moreover, this correlation remains significant when controlling for individual differences in other cognitive and performance factors. Our results show that individual differences in achievement in school mathematics are related to individual differences in the acuity of an evolutionarily ancient, unlearned approximate number sense. Further research will determine whether early differences in number sense acuity affect later maths learning, whether maths education enhances number sense acuity, and the extent to which tertiary factors can affect both.

Behavioural, neuropsychological and brain imaging techniques show that a signature of the approximate number system (ANS) is its imprecision<sup>2–13</sup>. Unlike exact verbal counting, the ANS produces numerical representations that grow increasingly imprecise as a linear function of the target array, with larger quantities represented less precisely than smaller quantities. This imprecision is expressed as a Weber fraction that indexes the amount of error in the underlying mental representation of any numerosity<sup>3–5</sup>. On average, the Weber fraction of adults is approximately 0.11, yielding successful non-verbal discrimination of arrays differing by as little as a 9:10 ratio<sup>5,14</sup>. Here we address whether there are significant individual differences in ANS acuity, and also whether these differences correlate with individual differences in symbolic maths achievement.

We examined 64 14-yr-old children with normal development whose performance in a variety of mathematical and more general cognitive tasks had been measured longitudinally, starting in kindergarten<sup>15</sup>. We tested for correlations between the current ANS acuity of the subjects and their past achievement in symbolic maths, while controlling for a wide range of other variables. Each subject’s ANS

acuity was assessed by psychophysical modelling of performance on a simple more/less judgement task similar to those used previously with infants and non-human animals. On each trial, subjects saw spatially intermixed blue and yellow dots presented on a computer screen too rapidly (200 ms) to serially count (Fig. 1a)<sup>16</sup>. Subjects indicated which colour was more numerous by key press and verbal response. The ratio between the two sets varied randomly among 1:2, 3:4, 5:6 and 7:8, with between 5 and 16 dots in each set. The colour of the more numerous set varied randomly, and half of the trials were area-controlled to ensure that responses were on the basis of the



**Figure 1 | Method and group performance.** **a**, A representation of the trial from the numerical discrimination task. **b**, Group performance and modelled best-fit for all trials in the numerical discrimination task. **c**, Histogram of  $w$ , the acuity of the ANS, for the sample ( $n = 64$ ), as determined by the psychophysical model for each subject.

<sup>1</sup>Johns Hopkins University, Ames Hall, 3400 North Charles Street, Baltimore, Maryland 21218, USA. <sup>2</sup>Kennedy Krieger Institute, 3825 Greenspring Avenue, Painter Building, Top Floor, Baltimore, Maryland 21211, USA.

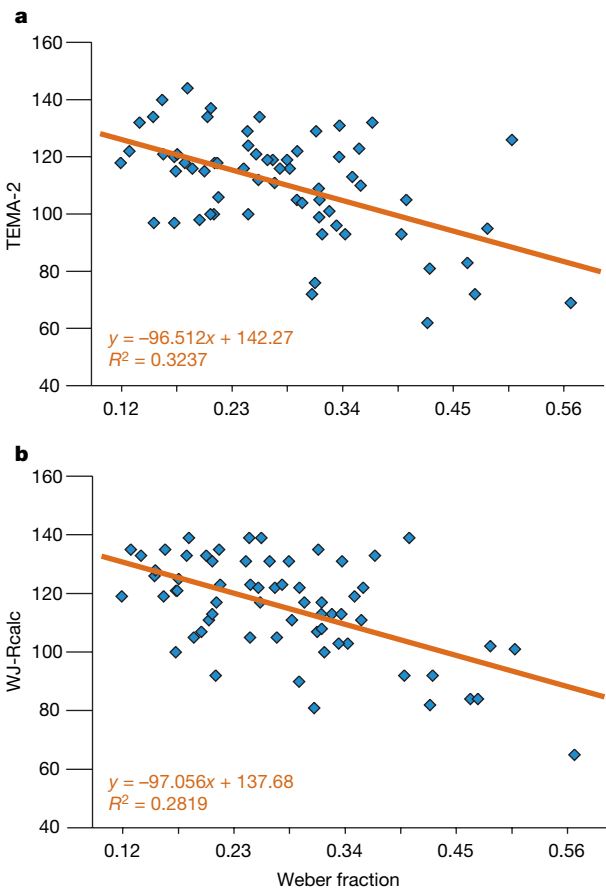


number of dots and not on the total dot area (see Supplementary Information). Subjects participated in two sessions of 10 practice trials and 40 test trials each, totalling 80 test trials (approximately 10 min of testing per subject).

Collapsing across subjects, numerical discrimination improved as the ratio between the presented numerosities increased, in accord with Weber’s law and with previous investigations of the ANS<sup>3–9</sup> (Fig. 1b). This gradual improvement in performance as a function of ratio was modelled using classical psychophysical tools to determine the group Weber fraction (see Methods and Supplementary Information). This returned a value of 0.265 for the group Weber fraction ( $w$ ) with an  $R^2$  value of 0.995, suggesting that there is very high agreement between this psychophysical model of the ANS and the behavioural data (Fig. 1b). Next, we used this same method to fit each individual subject’s data and thereby determine each subject’s Weber fraction. This showed surprisingly large variation in the ANS acuity ( $w$ ), ranging from 0.119 to 0.567 (Fig. 1c). The Weber fractions of subjects can also be translated into more intuitive whole numbers that show the ratio that would result in 75% correct performance. Using this translation, some subjects could discriminate numerical ratios as fine as 9:10 ( $w = 0.11$ ) whereas others had difficulty with ratios finer than 2:3 ( $w = 0.5$ ; mean subject  $w \approx 4:5$ ).

A question to address is whether these individual differences in ANS acuity ( $w$ ) predict individual differences in symbolic maths achievement. Each of our subjects was tested annually from kindergarten to sixth grade (ages 5–11) on a battery of standardized and investigator-designed measures. This longitudinal assessment of mathematical, verbal and other cognitive abilities provides a unique opportunity to detect any enduring correlations between ANS acuity and symbolic maths ability while controlling for other factors. Each year (ages 5–11), symbolic maths ability was assessed using the ‘test of early mathematical ability, second edition’ (TEMA-2)<sup>17</sup> and/or the ‘Woodcock–Johnson revised calculation subtest’ (WJ-Rcalc)<sup>18</sup>, yielding an age-referenced standardized score for each subject. We found that the ANS acuity ( $w$ ) of subjects correlated with symbolic maths performance in every year tested (from kindergarten to sixth grade) for both of the standardized maths tests, as summarized in Table 1. ANS acuity in ninth grade retrospectively predicted the symbolic maths performance of individual students from as early as kindergarten, a 9-yr time span. The linear correlations of ANS acuity ( $w$ ) with symbolic maths achievement (TEMA-2 and WJ-Rcalc) for the third grade are shown in Fig. 2a, b.

A further question to address was whether the correlation between ANS acuity and symbolic maths achievement was due to individual differences in more general cognitive or performance factors. In the third grade (when subjects were approximately aged 8) we administered several non-numerical standardized tests including measures of rapid lexical access for colour names (rapid automatic naming, RAN-colour)<sup>19</sup> and full-scale IQ (Wechsler abbreviated scale of intelligence, WASI-full)<sup>20</sup>. The RAN-colour is an appropriate control



**Figure 2 | Regressions.** **a**, **b**, Linear regression of the standard score for each subject on the TEMA-2 test (**a**) or on the WJ-Rcalc test (**b**) of symbolic maths achievement and the acuity of the ANS ( $w$ ). For TEMA-2 and WJ-Rcalc, higher numbers indicate better performance, whereas for the Weber fraction, lower numbers indicate better performance.

for our task because it measures the reaction time to identify the colours of 50 stimuli quickly; rapid colour naming is precisely the behaviour required by our ANS acuity assessment. The WASI-full IQ test acts as a control for general intelligence. WASI-full and RAN-colour did not correlate with one another in our sample ( $P = 0.699$ ), making them largely orthogonal for purposes of linear regressions with ANS acuity. To examine the relationship of ANS acuity and symbolic maths achievement while controlling for other variables, two separate linear regressions were performed with ANS acuity ( $w$ ) as the dependent variable and performance on either the TEMA-2 or the WJ-Rcalc test, and WASI-full and RAN-colour as independent variables. These showed that ANS acuity ( $w$ ) correlated with symbolic maths achievement in the third grade even with rapid lexical access and general intelligence controlled for (Table 2).

To assess the strength of the correlation between ANS acuity ( $w$ ) and symbolic maths achievement further, we performed extra linear regressions between  $w$  (measured at age 14) and an even broader range of standardized test scores obtained when subjects were in the third grade. These 16 measures controlled for the widest possible range of behavioural, cognitive and intelligence factors in our sample including many factors promoted as predictors of mathematical ability (for example, visual–spatial reasoning, working memory)<sup>21–25</sup>. ANS acuity ( $w$ ) significantly correlated with symbolic maths achievement (measured in the third grade) for both TEMA-2 and WJ-Rcalc performance, with all 16 measures controlled for ( $r_p^2 = 0.167$  and  $0.200$ , respectively, where  $p$  represents partial correlation). In contrast, no other measure correlated with ANS acuity when symbolic maths performance and other variables were controlled for (Table 3). This means that success on tests of symbolic mathematics throughout the school years

Table 1   Correlation of ANS acuity ( $w$ ) with symbolic maths achievement						
Grade	TEMA-2 $R^2$	$t$ d.f. = 62	$P$	WJ-Rcalc $R^2$	$t$ d.f. = 62	$P$
Kindergarten	0.137	3.134	0.003	0.127	2.959	0.004
First	0.140	3.171	0.002	0.326	5.480	$8 \times 10^{-7}$
Second	0.238	4.399	$4 \times 10^{-5}$	—	—	—
Third	0.324	5.448	$9 \times 10^{-7}$	0.282	4.933	$6 \times 10^{-6}$
Fourth	—	—	—	0.248	4.518	$3 \times 10^{-5}$
Fifth	—	—	—	0.117	2.866	0.006
Sixth	—	—	—	0.251	4.564	$2 \times 10^{-5}$

ANS acuity ( $w$ ) measured in ninth grade retroactively correlated with symbolic maths achievement.  $R^2$  values represent the proportion of the variance in symbolic maths achievement that is explained by ANS acuity.  $R^2$  values  $>0.25$  are considered large in behavioural science and are generally viewed as having large practical significance.  $t$  values represent the distance, measured in units of standard error, between the obtained correlation and the null hypothesis of no correlation.  $P$  values represent the probability of obtaining the observed correlation in a sample of data by random chance when there is truly no relation in the population.

**Table 2 | Correlations controlled for cognitive and performance factors**

Measure (task)	$r_p^2$	$t$	$P$
		d.f. = 60	
With TEMA-2			
Symbolic maths (TEMA-2)	0.146	3.205	0.002
Intelligence (WASI-full)	0.013	0.887	0.379
Task demands (RAN-colour)	0.004	0.492	0.625
With WJ-Rcalc			
Symbolic maths (WJ-Rcalc)	0.155	3.325	0.003
Intelligence (WASI-full)	0.070	2.124	0.038
Task demands (RAN-colour)	0.017	1.023	0.310

ANS acuity ( $w$ ) measured in ninth grade retroactively correlated with third grade symbolic maths achievement and other measures.  $r_p^2$  values represent the proportion of the variance in ANS acuity accounted for by the listed variable when controlling for the two remaining variables in each analysis (TEMA-2 or WJ-Rcalc).  $t$  values represent the distance, measured in units of standard error, between the obtained correlation and the null hypothesis of no correlation.  $P$  values represent the probability of obtaining the observed correlation in a sample of data by random chance when there is truly no relation in the population.

can be retrospectively predicted by a subject's ANS acuity in young adulthood, as measured by the simple task of determining which of two quickly flashed arrays has more dots, even with extensive controls for other cognitive and performance factors.

Our results are consistent with at least two interpretations. Given that it is functional in infancy<sup>7</sup>, long before the onset of symbolic mathematics instruction, the ANS may have a causal role in determining individual maths achievement. Indeed, neuropsychological evidence suggests that the ANS is activated during symbolic mathematical reasoning across the lifespan<sup>13</sup>; therefore individual differences in ANS acuity might give rise to individual differences in maths ability.

**Table 3 | Correlations controlled for all available factors**

Measure (task)	$r_p^2$	$t$	$P$
		d.f. = 40	
With TEMA-2			
Symbolic maths (TEMA-2)	0.167	2.831	0.007
Intelligence (WASI-full)	0.005	0.472	0.640
Task demands (RAN-colour)	0.023	0.981	0.332
Verbal IQ (WASI-verbal)	0.005	0.459	0.649
Performance IQ (WASI-performance)	0.006	0.482	0.632
Executive functions (CNT-B3)	0.021	0.918	0.364
Visual working memory (MemPuzl)	0.067	1.694	0.098
Visual segmentation (DTVPf)	0.009	0.599	0.552
Object perception (DTVPfc)	0.001	0.172	0.864
Visual reasoning (DTVPvc)	0.025	1.004	0.321
Spatial reasoning (DTVPps)	0.012	0.701	0.488
Visual motor integration (VMI)	0.035	1.213	0.232
Word knowledge (WJ-RIwid)	0.012	0.706	0.484
Reading (WJ-Rwa)	0.001	0.225	0.823
Rapid lexical access (RAN-letter)	0.049	1.435	0.149
Rapid lexical access (RAN-number)	0.012	0.685	0.497
Gender	0.028	1.069	0.291
With WJ-Rcalc			
Symbolic maths (WJ-Rcalc)	0.200	3.149	0.003
Intelligence (WASI-full)	0.013	0.736	0.466
Task demands (RAN-colour)	0.004	0.391	0.698
Verbal IQ (WASI-verbal)	0.009	0.605	0.549
Performance IQ (WASI-perf)	0.013	0.727	0.472
Executive functions (CNT-B3)	0.035	1.208	0.234
Visual working memory (MemPuzl)	0.084	1.916	0.062
Visual segmentation (DTVPf)	0.032	1.148	0.254
Object perception (DTVPfc)	0.001	0.201	0.842
Visual reasoning (DTVPvc)	0.008	0.578	0.566
Spatial reasoning (DTVPps)	0.018	0.869	0.390
Visual motor integration (VMI)	0.013	0.725	0.473
Word knowledge (WJ-RIwid)	0.014	0.757	0.454
Reading (WJ-Rwa)	0.000	0.014	0.988
Rapid lexical access (RAN-letter)	0.014	0.757	0.454
Rapid lexical access (RAN-number)	0.000	0.037	0.970
Gender	0.012	0.684	0.498

ANS acuity ( $w$ ) measured in ninth grade retroactively correlated with third grade symbolic maths achievement and other measures.  $r_p^2$  values represent the proportion of the variance in ANS acuity accounted for by the listed variable when controlling for all other variables in the list.  $t$  values represent the distance, measured in units of standard error, between the obtained correlation and the null hypothesis of no correlation.  $P$  values represent the probability of obtaining the observed correlation in a sample of data by random chance when there is truly no relation in the population.

Alternatively, individual differences in the quantity or quality of engagement in formal mathematics might increase ANS acuity. This latter possibility is hinted at by cross-cultural differences in Weber fractions, with maths-educated adults having better ANS acuity than adults from indigenous cultures lacking maths education<sup>5,14</sup>. These causal relationships, possible tertiary factors and the trainability of ANS acuity<sup>26</sup> remain to be explored. Further evidence will add to the present results, which suggest that our ability to reason over symbolic numbers is deeply entwined with an evolutionarily ancient system for numerical approximation.

## METHODS SUMMARY

At age 14 (that is, ninth grade), ANS acuity was assessed for 64 subjects (see Methods). The percentage correct on the ANS task was modelled for each individual subject as  $1 - \text{error rate}$ , where error rate is defined as:

$$\frac{1}{2} \operatorname{erfc} \left( \frac{n_1 - n_2}{\sqrt{2w} \sqrt{n_1^2 + n_2^2}} \right)$$

where  $\operatorname{erfc}(x)$  is the complementary error function related to the integration of the normalized Gaussian distribution. This model fits percentage correct as a function of the Gaussian approximate number representations for the two sets displayed on a trial ( $n_1$  and  $n_2$ , that is, blue dots and yellow dots) with a single free parameter, the Weber fraction ( $w$ ; see Supplementary Information)<sup>5</sup>. Correlations presented were between this estimate of ANS acuity ( $w$ ), measured at age 14, and scores on standardized cognitive and performance measures, from kindergarten to sixth grade.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 10 February; accepted 4 July 2008.

Published online 7 September 2008.

- Carey, S. Bootstrapping and the origin of concepts. *Daedalus* **133**, 59–68 (2004).
- Feigenson, L., Dehaene, S. & Spelke, E. S. Core systems of number. *Trends Cogn. Sci.* **8**, 307–314 (2004).
- Barth, H., Kanwisher, N. & Spelke, E. S. The construction of large number representations in adults. *Cognition* **86**, 201–221 (2003).
- Cordes, S., Gelman, R. & Gallistel, C. R. Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychon. Bull. Rev.* **8**, 698–707 (2001).
- Pica, P., Lemer, C., Izard, V. & Dehaene, S. Exact and approximate arithmetic in an Amazonian indigene group. *Science* **306**, 499–503 (2004).
- Gordon, P. Numerical cognition without words: Evidence from Amazonia. *Science* **306**, 496–499 (2004).
- Xu, F. & Spelke, E. S. Large number discrimination in 6-month-old infants. *Cognition* **74**, B1–B11 (2000).
- Nieder, A. & Miller, E. A. Parieto-frontal network for visual numerical information in the monkey. *Proc. Natl Acad. Sci. USA* **101**, 7457–7462 (2004).
- Meck, W. H. & Church, R. M. A mode control model of counting and timing processes. *J. Exp. Psychol. Anim. Behav. Process.* **9**, 320–334 (1983).
- Brannon, E. M. & Terrace, H. S. Ordering of the numerosities 1–9 by monkeys. *Science* **282**, 746–749 (1998).
- Dehaene, S., Dehaene-Lambertz, G. & Cohen, L. Abstract representations of numbers in the animal and human brain. *Trends Neurosci.* **21**, 355–361 (1998).
- Gallistel, C. R. *The Organization of Learning* (MIT Press, 1990).
- Dehaene, S. *The Number Sense: How the Mind Creates Mathematics*. (Oxford Univ. Press, 1997).
- Halberda, J. & Feigenson, L. Developmental change in the acuity of the “Number Sense”: The approximate number system in 3-, 4-, 5-, 6-year-olds and adults. *Dev. Psychol.* (in the press).
- Lachance, J. A. & Mazzocco, M. M. M. A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learn. Individ. Differ.* **16**, 195–216 (2006).
- Mandler, G. & Shebo, B. J. Subitizing: an analysis of its component processes. *J. Exp. Psychol. Gen.* **111**, 1–22 (1982).
- Ginsburg, H. & Baroody, A. *Test of Early Mathematics Ability* 2nd edn (Austin, 1990).
- Woodcock, R. W. & Johnson, M. B. *Woodcock-Johnson Psycho-Educational Battery-Revised* (DLM Teaching Resources, Allen, 1989).
- Denckla, M. B. & Rudel, R. G. Rapid automatized naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia* **14**, 471–479 (1976).
- Wechsler, D. *Wechsler Abbreviated Scale of Intelligence (WASI)* (Psychological Corporation, 1999).
- Fias, W. & Fischer, M. H. in *Handbook of Mathematical Cognition* (ed. Campbell, J.) 43–54 (Psychology Press, 2005).
- Mazzocco, M. M. M. & Myers, G. F. Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Ann. Dyslexia* **53**, 218–253 (2003).

23. Kurdek, L. A. & Sinclair, R. J. Predicting reading and mathematics achievement in fourth-grade children from kindergarten readiness scores. *J. Educ. Psychol.* **93**, 451–455 (2001).
24. Espy, K. A. *et al.* The contribution of executive functions to emergent mathematic skills in preschool children. *Dev. Neuropsychol.* **26**, 465–486 (2004).
25. Mazzocco, M. M. M. & Kover, S. T. A longitudinal assessment of executive function skills and their association with math performance. *Child Neuropsychol.* **13**, 18–45 (2007).
26. Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L. & Dehaene, S. An open trial assessment of “The Number Race”, an adaptive computer game for remediation of dyscalculia. *Behav. Brain Funct.* **2**, 19 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Longitudinal data collection was supported by a National Institutes of Health grant RO1 HD 034061 to M.M. We thank the Baltimore County Public School District and the children and parents who participated in this research, D. Naimen and R. Goodman for help with modelling, and L. Stapleton for help with statistical analyses.

**Author Contributions** J.H., M.M. and L.F. conceived the experiment; J.H. designed the numerical discrimination procedure; M.M. provided longitudinal data and oversaw data collection; J.H. performed the modelling and data analysis; J.H., L.F. and M.M. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.H. ([halberda@jhu.edu](mailto:halberda@jhu.edu)).



## METHODS

**Subjects.** Sixty-four 14-yr-old children participated (32 male; mean age 14 yr 10 months, ranging from 14 yr 3 months to 15 yr 9 months). They were from lower-middle to upper-middle economic backgrounds and were enrolled in normal (that is, non-learning-disabled) public classrooms in a single suburban school district outside of Baltimore, Maryland, USA. Throughout the longitudinal portion of the study, subjects were tested yearly in one, two or three 1-h sessions. The numerical discrimination task was performed during a laboratory visit during the tenth year of the longitudinal study (that is, ninth grade). Further details are given in the Supplementary Information. A total of 80 14-yr-olds were tested during the tenth year assessment but 9 were removed from the final sample owing to high variability in their performance on the ANS acuity assessment (see Modelling and analysis section), and 7 were removed because of missing data for some portion of the standardized tests in grades kindergarden to sixth grade.

**Numerical discrimination.** Subjects completed this task twice at an interval of approximately 60 min, as the first and last sub-tasks in a larger test battery from the longitudinal study. Each run of the task lasted 5 min. Subjects viewed dot arrays on a computer screen and judged whether there were more blue or more yellow dots. For each trial, pressing the space bar initiated a 250 ms blank-screen delay followed by a 200 ms appearance of an array of intermixed blue and yellow dots. After the array had disappeared, subjects had an unlimited amount of time to indicate their response by pressing a colour-coded keyboard button and saying the name of the more numerous colour aloud. Reaction time averaged approximately 1,100 ms across subjects. Subjects were told that, if they wished to change their choice, they could correct an erroneous key-press response by reporting the intended response to the experimenter, who noted it on a score sheet. Self corrections were reported by only 6 of the 64 subjects and accounted for only 7 out of the 5,120 total trials recorded in the study. The number of dots in each set in the array ranged from 5 to 16. Whether the yellow or blue set was larger was randomized. Each trial was drawn from one of four ratio bins in which the ratio of the smaller to the larger set was 1:2, 3:4, 5:6 or 7:8. For each of two runs of the experiment, subjects received 10 practice trials randomly selected from these ratios followed by 40 randomly ordered test trials (10 trials per ratio). Half of the trials in each ratio were 'dot-size controlled': the size of the average blue dot was equal to the size of the average yellow dot. On these trials, the set with more dots necessarily also had a larger total area on screen. The other half of trials were 'area controlled': the total number of blue pixels equalled the total number of yellow pixels such that the total cumulative area of the two sets was identical. The set with more dots thereby had smaller dots on average. Because the two sets were spatially overlapping and each dot was randomly placed in a shared display window, area-controlled trials also controlled for other continuous variables associated with number such as total dot density, inter-dot distance and the total envelope size of each set. Preliminary analyses showed similar results for dot-size-controlled and area-controlled trials and these data were

therefore combined for each subject. On both dot-size-controlled and area-controlled trials, individual dot size varied randomly by up to  $\pm 35\%$  of the set average to discourage the use of individual dot size as a proxy for number. The diameter of a typical dot subtended approximately 1 degree of visual angle from a viewing distance of 50 cm.

**Modelling and analysis.** Previous investigations have modelled numerical representations either as having linearly increasing means and linearly increasing standard deviation<sup>27</sup>, or as having logarithmically compressed means with constant standard deviation<sup>8</sup>. Both of these formats capture the performance pattern that is characteristic of the ANS (error that increases linearly with target numerosity). We used a classical psychophysics model that relies on a linear format of the ANS (although a logarithmic model makes the same predictions for our simple numerical acuity task), which provides a psychologically plausible model of performance in numerical discrimination<sup>5</sup>. Percentage correct was modelled as a function of increasing ratio (larger set/smaller set, or  $n_2/n_1$ ). The numerosity for the blue set and yellow set were represented as Gaussian random variables (that is,  $X_2$  and  $X_1$ ) with means  $n_2$  and  $n_1$  and standard deviations equal to the Weber fraction  $w \times n$ . Subtracting the Gaussian for the smaller set from the larger set returned a new Gaussian with a mean of  $n_2 - n_1$  and a standard deviation of  $w\sqrt{n_1^2 + n_2^2}$  (simply the difference of two Gaussian random variables). Percentage correct was then equal to  $1 - \text{error rate}$ , in which error rate is defined as the area under the tail of the resulting Gaussian curve, computed as:

$$\frac{1}{2} \operatorname{erfc} \left( \frac{n_1 - n_2}{\sqrt{2} w \sqrt{n_1^2 + n_2^2}} \right)$$

This model fits percentage correct on the numerical discrimination task as a function of the Gaussian approximate number representations for the two sets (that is, blue and yellow dots) with a single free parameter, the Weber fraction ( $w$ ). An individual subject's Weber fraction ( $w$ ) describes the standard deviations for the Gaussian representations of the ANS, thereby describing the amount of overlap between any two Gaussian representations, and thereby predicting percentage correct for any numerical discrimination. Using this model, the best-fit value for the Weber fraction ( $w$ ) was determined by a program implementing the Levenberg–Marquardt algorithm for nonlinear least-squares fit on the average percentage correct in each ratio bin for each subject. The model attempts to determine the best-fit value for  $w$  in 50 iterations, each iteration being an attempt to reduce the sum of squared error. The model did not settle on a value for 9 of the original 80 subjects we tested, owing to high variability in the accuracy of their responses. These subjects were removed from the analysis.

27. Brannon, E. M., Wusthoff, C. J., Gallistel, C. R. & Gibbon, J. Numerical subtraction in the pigeon: evidence for a linear subjective number scale. *Psychol. Sci.* **12**, 238–243 (2001).

# UNC-6/netrin and its receptor UNC-5 locally exclude presynaptic components from dendrites

Vivian Y. Poon<sup>1</sup>, Matthew P. Klassen<sup>1</sup> & Kang Shen<sup>1,2</sup>

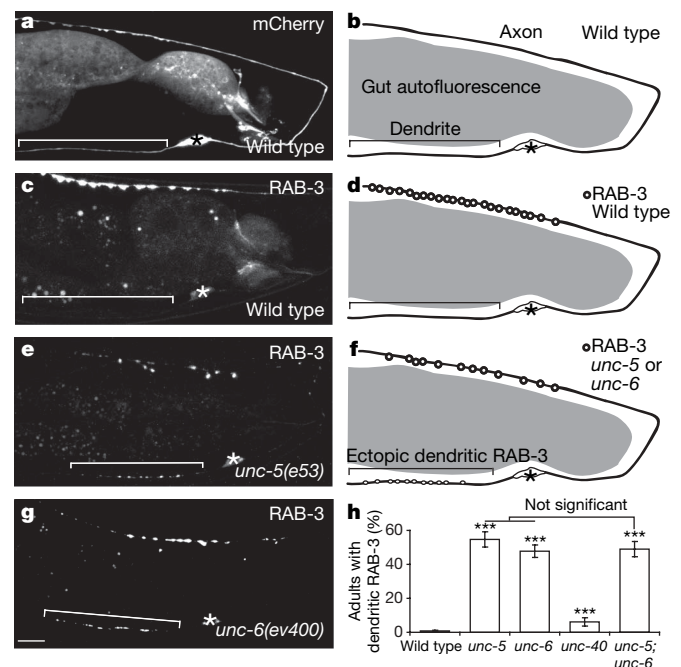
Polarity is an essential feature of many cell types, including neurons that receive information from local inputs within their dendrites and propagate nerve impulses to distant targets through a single axon. It is generally believed that intrinsic structural differences between axons and dendrites dictate the polarized localization of axonal and dendritic proteins<sup>1</sup>. However, whether extracellular cues also instruct this process *in vivo* has not been explored. Here we show that the axon guidance cue UNC-6/netrin and its receptor UNC-5 act throughout development to exclude synaptic vesicle and active zone proteins from the dendrite of the *Caenorhabditis elegans* motor neuron DA9, which is proximal to a source of UNC-6/netrin. In *unc-6/netrin* and *unc-5* loss-of-function mutants, presynaptic components mislocalize to the DA9 dendrite. In addition, ectopically expressed UNC-6/netrin, acting through UNC-5, is sufficient to exclude endogenous synapses from adjacent subcellular domains within the DA9 axon. Furthermore, this anti-synaptogenic activity is interchangeable with that of LIN-44/Wnt despite being transduced through different receptors, suggesting that extracellular cues such as netrin and Wnts not only guide axon navigation but also regulate the polarized accumulation of presynaptic components through local exclusion.

The *C. elegans* motor neuron DA9 elaborates a molecularly and functionally distinct axon and dendrite<sup>2</sup> (Fig. 1a, b). In wild-type animals, presynaptic components are excluded from the dendrite and accumulate in a stereotyped and discrete domain within the DA9 dorsal axon<sup>3</sup> (Fig. 1c, d). These presynaptic components include synaptic vesicle proteins such as RAB-3, SNB-1/synaptobrevin and SNG-1/synaptogyrin (Supplementary Fig. 1), the L-type voltage-gated calcium channel  $\beta$ -subunit CCB-1, and the active zone protein SYD-2/ $\alpha$ -liprin (Supplementary Fig. 2). In exploring whether extracellular cues instruct this polarized localization, we found that these presynaptic proteins mislocalize to the DA9 dendrite in *unc-6/netrin* (*ev400*) and *unc-5(e53)* null mutants (Fig. 1e–h, and Supplementary Figs 1 and 2). This mislocalization defect is not enhanced in *unc-5;unc-6/netrin* double mutants, suggesting that UNC-5 and UNC-6/netrin function in the same pathway (Fig. 1h). A null mutation in the other principal UNC-6/netrin receptor, UNC-40, results in a minor mislocalization defect (Fig. 1h). We further observed that this mislocalization is partly suppressed by a mutation in the presynaptic assembly gene, *syd-2/liprin- $\alpha$*  (Supplementary Fig. 3), suggesting that SYD-2/ $\alpha$ -liprin promotes the accumulation of GFP::RAB-3 in the DA9 dendrite. In addition to the mislocalization defect, the average number of GFP::RAB-3 puncta in the dorsal axon of DA9 is reduced in *unc-5* mutants compared with wild-type animals (Fig. 1e–g and Supplementary Fig. 4).

Netrins are evolutionarily conserved axon guidance molecules present in worms<sup>4</sup>, flies<sup>5</sup> and mammals<sup>6</sup>. The activity of these secreted

molecules is transmitted through two distinct cell-surface receptors: UNC-5 repels axons<sup>7</sup> whereas UNC-40/DCC/Frazzled<sup>8–10</sup> attracts axons to a source of UNC-6/netrin. Similar to mammals, UNC-6/netrin in *C. elegans* is expressed in many classes of ventral cells and its expression persists into adulthood<sup>4</sup>. The UNC-5 receptor is expressed in DA motor neurons, as indicated by antibody staining, transgene expression<sup>11</sup> and microarray analysis<sup>12</sup>. We confirmed that UNC-5 is expressed in DA9 with a transgenic line expressing dsRed driven by the *unc-5* promoter (Supplementary Fig. 5a–c).

To distinguish if UNC-5 is required for localization of presynaptic components in DA9 itself or in other cells like the postsynaptic partners of DA9 (VD/DD neurons and dorsal body-wall muscles), we performed cell-autonomous rescue experiments with various promoters. The *mig-13* promoter is expressed only in DA9 within the tail region at the early larval L1 stage<sup>13</sup> and a *mig-13::unc-5* transgene



**Figure 1 | GFP::RAB-3 is mislocalized to the dendrite in *unc-5* and *unc-6/netrin* mutants.** a–d, Micrographs and diagrams of representative wild-type adults expressing cytoplasmic mCherry (a, b) or GFP::RAB-3 (c, d). e–g, Micrographs and diagram of representative *unc-5* and *unc-6/netrin* mutant adults expressing GFP::RAB-3. Signal in the middle of the worm is gut autofluorescence. Anterior, left; dorsal, top; brackets, dendrites; asterisks, cell bodies. Scale bar, 10  $\mu$ m. h, Penetrance of dendritic GFP::RAB-3 in adults with no DA9 guidance defects. Error bars, standard error of proportion;  $n > 100$ ; \*\*\* $P < 0.0001$  (versus wild-type animals),  $\chi^2$  test.

<sup>1</sup>Neuroscience Program, Stanford University School of Medicine, 300 Pasteur Drive, California 94305, USA. <sup>2</sup>Department of Biology, Howard Hughes Medical Institute, Stanford University, 385 Serra Mall, California 94305, USA.

robustly rescues the mislocalization defect in *unc-5* mutant L1 and adult animals, suggesting that UNC-5 acts cell-autonomously in DA9 to exclude presynaptic components from the dendrite. Furthermore, we did not observe any rescue when UNC-5 was expressed in the postsynaptic partners of DA9 using the *unc-25* and *unc-129m* promoters<sup>14,15</sup> (Fig. 2a and Supplementary Fig. 6). To substantiate these observations, we created *unc-5* mutant animals expressing a rescuing *unc-5::unc-5* transgene together with a cytoplasmic DA9 marker in a mosaic pattern. In two independent transgenic lines, we observed a strong correlation between the expression of *unc-5* in DA9 and rescue of the mislocalization defect, consistent with a cell-autonomous function for UNC-5 in DA9 (Fig. 2b). Using an UNC-5::YFP fusion construct expressed specifically in DA9, we observed a higher level of UNC-5::YFP in the dendrite and ventral axon than in the dorsal axon (Supplementary Fig. 7).

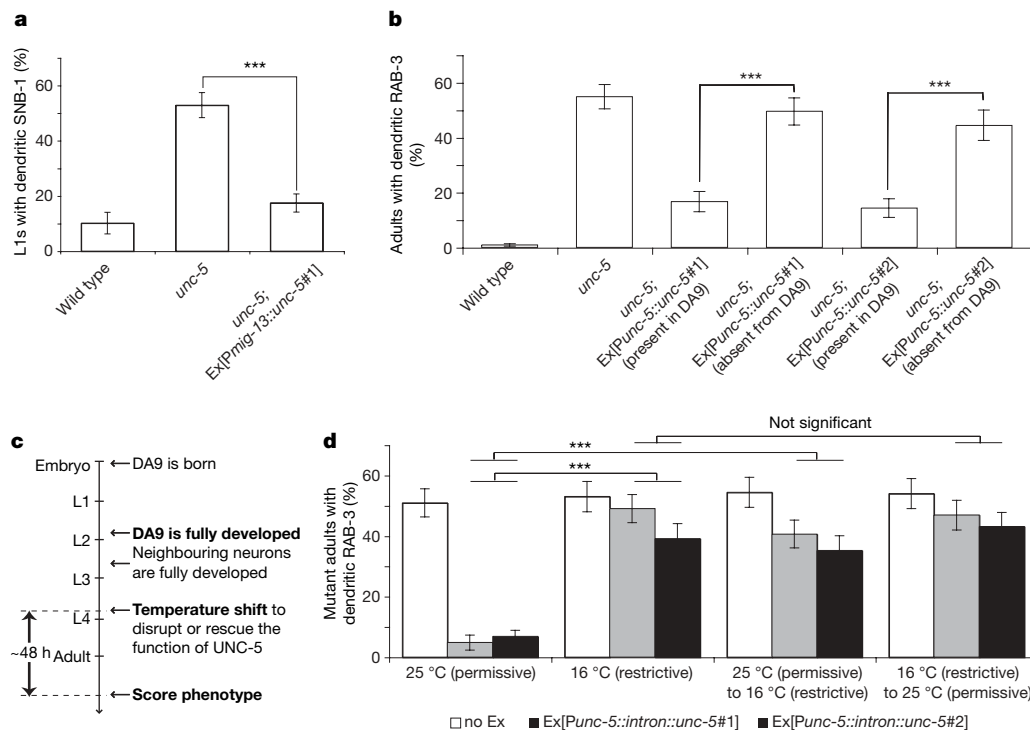
The disrupted distribution of presynaptic components in *unc-5* and *unc-6/netrin* mutants is observable at the early L1 larval stage, when the DA9 dendrite begins to form, and persists throughout the life of the animal (Supplementary Figs 1 and 2a–f). To differentiate whether the mislocalization of presynaptic components is a consequence of an early axodendritic polarity or guidance defect or a later developmental defect, we used a modified version of a silencing intron cassette (M. Chalfie, personal communication) to regulate *unc-5* temporally. We observed that culturing *unc-5* mutant animals expressing the *unc-5::intron::unc-5* transgene at 25 °C throughout development resulted in a significant rescue of the mislocalization defect in *unc-5* mutants, whereas culturing them at 16 °C led to no rescue, suggesting that the transgene produces functional UNC-5 at 25 °C but not at 16 °C (Fig. 2d and Supplementary Fig. 8a). A shift to the restrictive temperature at the L4 larval stage, after DA9 and surrounding neurons are fully developed, results in a mislocalization defect in the transgenic mutant animals that is comparable to *unc-5* mutants. Conversely, a shift to the permissive temperature is insufficient to rescue the mislocalization defect, suggesting that the defect

is irreversible. Therefore, compromising the activity of *unc-5* in mature DA9 neurons leads to a mislocalization defect, suggesting a novel function for UNC-5 in maintaining the polarized localization of GFP::RAB-3 independent of early polarization and guidance.

To elucidate further whether this novel function of UNC-5 can be separated from its previously known role in axon guidance, we examined axon guidance in the DD and VD neurons using the same temperature-shift experimental paradigm. We found that a shift to the restrictive temperature at the L4 larval stage did not cause further errors in axon guidance, suggesting that UNC-5 is only required during the early outgrowth phase to guide axons, and is not required later to maintain axon trajectory (Supplementary Fig. 8b).

The two distinct roles of UNC-5 in axon guidance and GFP::RAB-3 localization is further supported by the following observations. First, approximately half of *unc-5* or *unc-6/netrin* mutant animals have no detectable DA9 guidance defects, yet the mislocalization defect is still observed (Supplementary Figs 9 and 10a). Second, *unc-5* and *unc-6/netrin* mutant animals with defective DA9 guidance do not display more penetrant mislocalization defects (Supplementary Fig. 10a). Third, UNC-129/TGF- $\beta$ , like UNC-5 and UNC-6/netrin, is important for the dorsal guidance of DA neurons<sup>15</sup>, and we observed that approximately half of *unc-129/Tgf- $\beta$*  mutant animals have severe DA9 guidance defects but none exhibit dendritic GFP::RAB-3 (Supplementary Fig. 10b, c).

What cell biological processes do UNC-5 and UNC-6/netrin affect in localizing presynaptic components? It is possible that they are important for the establishment and maintenance of dendritic fate. Hence, we examined the localization of four dendritically localized proteins: CAM-1/ROR<sup>16</sup>, a receptor tyrosine kinase that localizes somatodendritically in hippocampal cultures<sup>17</sup>; UNC-9/innexin, a structural component of invertebrate gap junctions<sup>18</sup>; F35D2.3/fibrillin<sup>16</sup>; and DYS-1/dystrophin<sup>16</sup>. In wild-type animals, CAM-1, UNC-9 and DYS-1 localize to the DA9 dendrite, cell body and ventral axon, whereas F35D2.3/fibrillin localizes exclusively to the dendrite. These



**Figure 2 | UNC-5 acts cell autonomously in DA9 and is required throughout development.** **a**, Expression of UNC-5 in DA9 rescues the mislocalization defect in *unc-5* mutant L1 animals ( $n > 100$ ). **b**, Expression of UNC-5 in other cells does not rescue the mislocalization defect ( $n > 80$ ). **c**, Experimental timeline. **d**, The *unc-5::intron::unc-5* transgene rescues the

mislocalization defect of *unc-5;mec-8* mutant adults at 25 °C, not 16 °C. The mislocalization defect occurs when UNC-5 is inactivated early or late in development. Error bars, standard error of proportion;  $n > 100$ ; \*\*\* $p < 0.0001$  (within each transgenic line),  $\chi^2$  test.



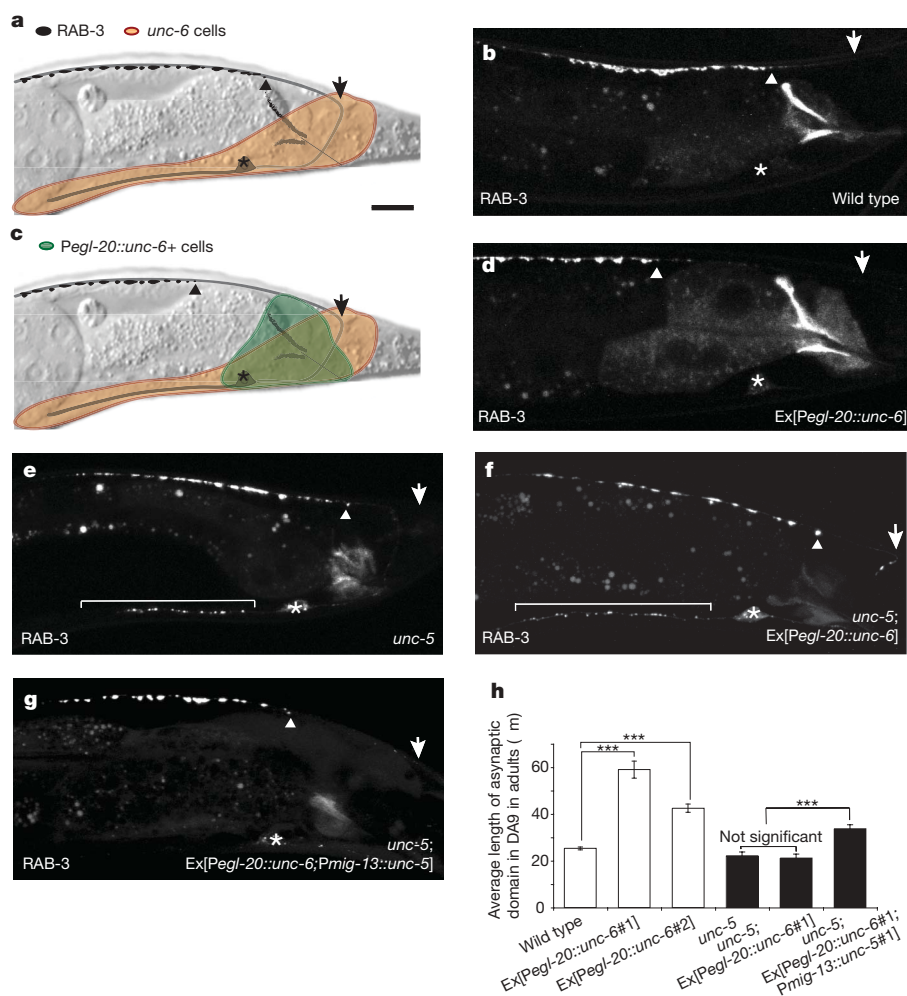
localization patterns are unaffected in *unc-5* and *unc-6/netrin* mutants, suggesting that many aspects of axodendritic polarization are maintained in these mutants (Supplementary Fig. 11). These results are consistent with the late temporal requirement of UNC-5 for proper localization of presynaptic components.

An alternative possibility is that UNC-5 and UNC-6/netrin regulate accumulation of presynaptic components in the dorsal axon, and that reduced axonal accumulation may indirectly cause the mislocalization defect. However, we did not observe any significant difference when we compared the average number of axonal GFP::RAB-3 puncta between *unc-5* mutants with and without dendritic GFP::RAB-3 (Supplementary Fig. 3). We conclude that this model is unlikely to be true.

To test directly whether UNC-6/netrin provides instructive information for the localization of presynaptic components, we generated a posterior to anterior gradient of UNC-6/netrin near DA9 using the *egl-20* promoter<sup>19</sup> where the posterior segment of the DA9 dorsal axon is exposed to an abnormally high level of UNC-6/netrin. This ectopic expression of UNC-6/netrin causes DA9 guidance defects in a small fraction of animals ( $12.2 \pm 3.2\%$ ). When we examined animals with normal DA9 guidance, we observed that the *egl-20::unc-6/netrin* transgene dramatically displaces GFP::RAB-3 anteriorly, creating an enlarged asynaptic zone in the posterior segment of the DA9 dorsal axon compared with wild-type animals (Fig. 3a–d, h). As the transgene does not affect the dorsal

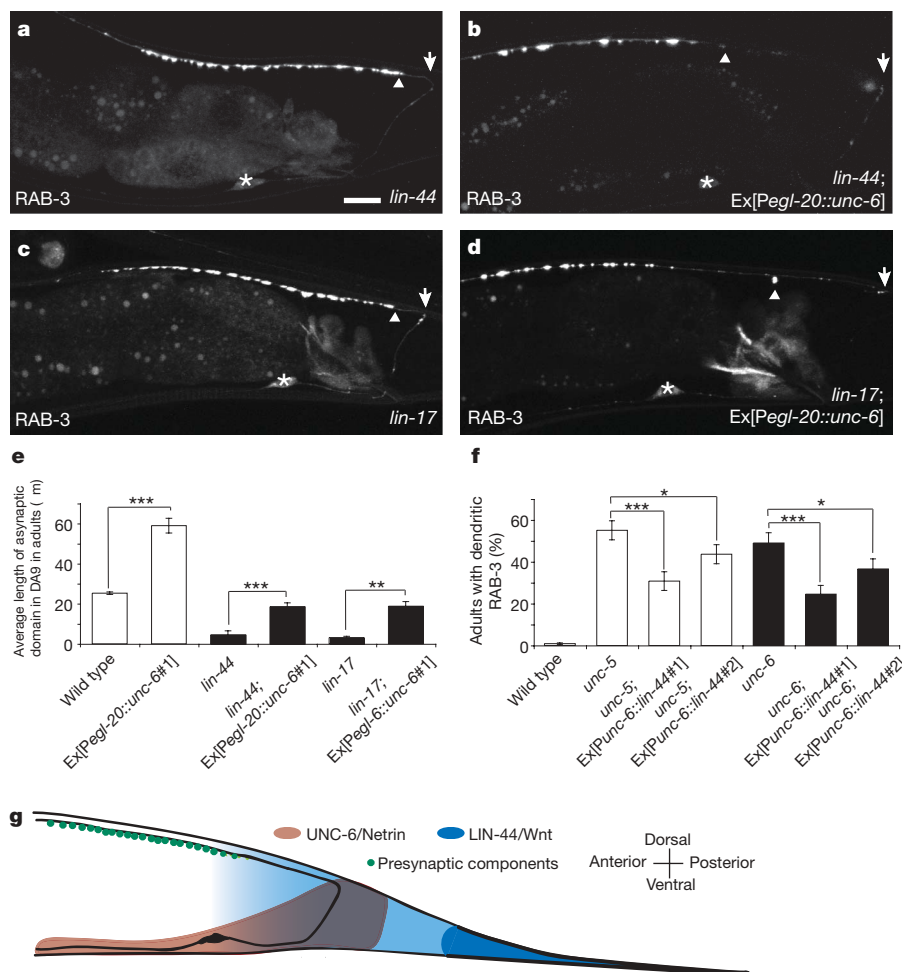
axon length of DA9 (Supplementary Fig. 12), it is unlikely that altered axonal outgrowth of DA9 leads to the enlarged asynaptic zone. We further observed that *unc-5* mutants expressing the *egl-20::unc-6/netrin* transgene do not have an enlarged asynaptic domain (Fig. 3e, f). In addition, the enlarged asynaptic domain is partly restored in *unc-5* mutants expressing the *mig-13::unc-5* transgene, demonstrating that UNC-5 acts cell-autonomously in DA9 to mediate ectopic UNC-6/netrin-induced exclusion of presynaptic components (Fig. 3g, h).

The striking similarity between the displacement of presynaptic components induced by the *egl-20::unc-6/netrin* and *egl-20::lin-44/wnt* transgenes<sup>3</sup> suggests that UNC-6/netrin and LIN-44/Wnt can both exclude synapses. We observed that the asynaptic domains in *lin-44/wnt(n1792);Ex[Pegl-20::unc-6/netrin]* or *lin-17/fz(n671);Ex[Pegl-20::unc-6/netrin]* mutants are significantly larger than those in *lin-44/wnt* or *lin-17/fz* mutants alone, suggesting that ectopically expressed UNC-6/netrin is sufficient to rescue the mislocalization defect in *lin-44/wnt* and *lin-17/fz* mutants (Fig. 4a–e). We further observed a significant reduction in penetrance of the mislocalization defects in *unc-5* and *unc-6/netrin* mutants expressing *lin-44/wnt* under the control of the *unc-6* promoter (Fig. 4f). If UNC-6/netrin and LIN-44/Wnt have similar functions, one might expect that *unc-6/netrin;lin-44/wnt* double mutants would have more severe mislocalization defects. However, these double mutants exhibit a fully penetrant guidance defect such that the DA9 axon turns posteriorly (data not shown), precluding analysis of presynaptic



**Figure 3 | UNC-6/netrin is sufficient to exclude GFP::RAB-3 locally and acts through UNC-5.** **a**, Endogenous UNC-6/netrin-expressing cells (beige). **c**, Ectopically expressed UNC-6/netrin (green). **b, d**, Representative wild-type L4 animal expressing GFP::RAB-3 in the absence (**b**) or presence (**d**) of ectopic UNC-6/netrin. **e–g**, Representative *unc-5* mutant adults expressing

GFP::RAB-3 with (**f**) or without (**e**) ectopic UNC-6/netrin or, alternatively, ectopic UNC-6/netrin with the *mig-13::unc-5* transgene (**g**). Arrows, posterior end of dorsal axon; arrowheads, posterior border of presynaptic domain; brackets, dendrites; asterisks, cell bodies. Scale bar, 10 μm. **h**, Error bars, s.e.m. ( $n > 50$ ); \*\*\* $P < 0.0001$ ,  $t$ -test.



**Figure 4 | UNC-6/netrin and LIN-44/Wnt function interchangeably.**

**a–d**, Representative *lin-44/wnt* or *lin-17/fz* mutant adults expressing GFP::RAB-3 with (**b, d**) or without (**a, c**) ectopic UNC-6/netrin. Arrows, posterior end of dorsal axon; arrowheads, posterior border of presynaptic domain; brackets, dendrites; asterisks, cell bodies. Scale bar, 10 μm. **e**, Error

bars, s.e.m. ( $n > 50$ ). \*\*\* $P < 0.0001$ ,  $t$ -test. **f**, Error bars, standard error of proportion ( $n > 100$ ); \*\*\* $P < 0.0001$ ; \*\* $P < 0.005$ ; \* $P < 0.05$ ;  $\chi^2$  test. **g**, Model for the roles of UNC-6/netrin and LIN-44/Wnt in subcellular patterning of presynaptic specializations in DA9.

localization. Collectively, these results suggest that UNC-6/netrin and LIN-44/Wnt play parallel roles in specifying the discrete presynaptic domain of DA9 by excluding presynaptic components from inappropriate compartments (Fig. 4g). Ventrally secreted UNC-6/netrin excludes presynaptic components from the DA9 dendrite, whereas LIN-44/Wnt secreted by the tail hypodermal cells performs a similar function in the posterior segment of the DA9 dorsal axon.

Here we demonstrate a novel role for UNC-6/netrin in providing spatial information for the exclusion of presynaptic components throughout development. Interestingly, UNC-6/netrin was recently shown to promote presynaptic formation in the amphid interneuron AIY in *C. elegans*<sup>20</sup>. These opposing effects of UNC-6/netrin on presynaptic formation might be explained by the different receptors used: UNC-5 in DA9 and UNC-40/DCC in AIY. The roles of these receptors in synaptic polarization parallel their contrasting functions in axon guidance, with UNC-5 functioning in repulsion and UNC-40/DCC in attraction<sup>7–10</sup>.

In addition to its well-characterized function in axon guidance, UNC-6/netrin was recently implicated in the initial polarization of the *C. elegans* hermaphrodite-specific neuron (HSN) neuronal cell body<sup>21</sup>. However, it is unclear whether UNC-6/netrin is required for later stages of neuronal polarity. Our findings suggest that UNC-6/netrin and UNC-5 activity coordinate two temporally distinct functions in DA9: axons are first guided to the appropriate locations, and presynaptic components are later localized in a polarized manner.

Consistent with this hypothesis, netrin is expressed in the adult mammalian nervous system long after axon guidance is complete<sup>22</sup>.

The conventional view of synapse formation is that contact between synaptic partners triggers assembly of the pre- and post-synaptic apparatus through the interaction of adhesion molecules like neuroligin/neurexin, SynCAM and EphrinB/EphB receptor<sup>23</sup> across the synaptic cleft. However, extracellular cues such as members of the Wnt<sup>24,25</sup>, fibroblast growth factor<sup>26</sup> and bone morphogen protein<sup>27</sup> families can also promote synapse formation. Our studies in DA9 suggest that negative regulators also pattern synaptogenesis by inhibiting the accumulation of presynaptic components in inappropriate subcellular domains. The UNC-6/netrin gradient is high ventrally and low dorsally<sup>4</sup>, encompassing the dendrite and ventral axon of DA9. The LIN-44/Wnt gradient is high posteriorly and low anteriorly<sup>28</sup>, effectively reaching the ventral axon, commissure and posterior region of the DA9 dorsal axon. Signalling through independent receptors, both UNC-6/netrin and LIN-44/Wnt, excludes presynaptic components, setting negative constraints on presynaptic formation in DA9 and forcing synapses to form in a discrete domain within the DA9 dorsal axon (Fig. 4g). Thus, inhibitory factors play essential roles in patterning the subcellular distribution of synapses.

## METHODS SUMMARY

**Strains and genetics.** Worms were raised on OP50 *Escherichia coli*-seeded nematode growth medium (NGM) plates at 22 °C. The following mutant strains were

obtained through the Caenorhabditis Genetics Center: CB271 *unc-40(e271)*I, MT1306 *lin-17(n671)*I, MT4051 *lin-44(n1792)*I, NW987 *unc-129(ev554)*IV, MT464 *unc-5(e53)*IV;*dpy-11(e224)*V;*lon-2(e678)*X, NW434 *unc-6(ev400)*X, CB620 *lin-18(e620)*X and CZ900 *syd-2(ju37)*X. TU218 *mec-8(u218)*I was provided by M. Chalfie. N2 Bristol was used as the wild-type reference strain.

**Cloning and constructs.** Expression clones were made in a derivative of pPD49.26 (A. Fire), the pSM vector (S. McCarroll and C. I. Bargmann, personal communication). The plasmids and strains *wyIs85(Pitr-1 pB::gfp::rab-3)*, *wyIs92(Pmig-13::snb-1::yfp)*, *wyEx403(Pitr-1 pB::cam-1::yfp)*, *wyEx771(Pmig-13::gfp::ccb-1; Pmig-13::mcherry::rab-3)* and *wyIs75(Punc-47::rfp; Pexp-1::gfp)* (G. Maro and K.S., unpublished observations) were generated as previously described<sup>3</sup>. The following plasmids and strains were generated using standard techniques<sup>29</sup>: *wyEx1902(Pitr-1 pB::mcherry)*, *wyEx2055(Pitr-1 pB::gfp::syd-2; Pitr-1 pB::mcherry::rab-3)*, *wyIs109(Pmig-13::cfp::rab-3; Pmig-13::snb-1::yfp; Pmig-13::sng-1::mcherry)*, *wyEx1054(Pitr-1 pB::unc-9::yfp)*, *wyEx2396(Pitr-1 pB::F35D2.3/fibrillin::yfp)*, *wyEx2430(Pitr-1 pB::dys-1::yfp)*, *wyEx1311(Pmig-13::unc-5#1)*, *wyEx1485(Pmig-13::unc-5#2)*, *wyEx1498(Punc-5::unc-5::sl2::dsred)*, *wyEx1228(Punc-5::unc-5)*, *wyEx2419(Pmig-13::mcherry; Punc-5::unc-5#1)*, *wyEx2418(Pmig-13::mcherry; Punc-5::unc-5#2)*, *wyEx1277(Pitr-1 pB::unc-5::yfp)*, *wyEx2306(Punc-5::intron::unc-5#1)*, *wyEx2308(Punc-5::intron::unc-5#2)*, *wyEx1904(Pegl-20::unc-6#1)*, *wyEx1916(Pegl-20::unc-6#2)*, *wyEx2093(Punc-6::lin-44#1)* and *wyEx2094(Punc-6::lin-44#2)*. We used the co-injection markers *Podr-1::GFP* or *dsRED* injected at 20 ng  $\mu$ l<sup>-1</sup> or *Pttx-3::CFP* injected at 50 ng  $\mu$ l<sup>-1</sup>; transgenes were introduced into *C. elegans* as previously described<sup>29</sup>.

**Fluorescence microscopy and confocal imaging.** Images of fluorescently tagged fusion proteins were captured in live *C. elegans* using a Plan-Apochromat 63 $\times$ /1.4 objective on a Zeiss LSM510 confocal microscope. Worms were immobilized using 10 mM levamisole (Sigma). Only animals with no detectable DA9 guidance defects were used for imaging and quantification, unless otherwise specified.

**Quantification of average synaptic domain or dorsal axon length.** Asynaptic domain length was quantified as previously described<sup>3</sup>. Animals in the mid-L4 larval stage expressing a DA9-specific cytoplasmic marker (*wyEx1902*) were imaged using a Zeiss Axiophot with AxioCam/Axiovision digital imaging. As the DA9 dorsal axon extends slightly anterior to the vulva, the distance between the end of the axon and the vulva, a landmark for the mid-body, was measured.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 29 May; accepted 29 July 2008.

Published online 7 September 2008; corrected 2 October 2008 (details online).

- Horton, A. C. & Ehlers, M. D. Neuronal polarity and trafficking. *Neuron* **40**, 277–295 (2003).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the ventral nerve cord of *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **275**, 327–348 (1976).
- Klassen, M. P. & Shen, K. Wnt signaling positions neuromuscular connectivity by inhibiting synapse formation in *C. elegans*. *Cell* **130**, 704–716 (2007).
- Wadsworth, W. G., Bhatt, H. & Hedgecock, E. M. Neuroglia and pioneer neurons express UNC-6 to provide global and local netrin cues for guiding migrations in *C. elegans*. *Neuron* **16**, 35–46 (1996).
- Mitchell, K. J. *et al.* Genetic analysis of Netrin genes in *Drosophila*: Netrins guide CNS commissural axons and peripheral motor axons. *Neuron* **17**, 203–215 (1996).
- Serafini, T. *et al.* Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell* **87**, 1001–1014 (1996).
- Leung-Hagstjorn, C. *et al.* UNC-5, a transmembrane protein with immunoglobulin and thrombospondin type 1 domains, guides cell and pioneer axon migrations in *C. elegans*. *Cell* **71**, 289–299 (1992).
- Chan, S. S. *et al.* UNC-40, a *C. elegans* homolog of DCC (Deleted in Colorectal Cancer), is required in motile cells responding to UNC-6 netrin cues. *Cell* **87**, 187–195 (1996).
- Keino-Masu, K. *et al.* Deleted in Colorectal Cancer (DCC) encodes a netrin receptor. *Cell* **87**, 175–185 (1996).

- Kolodziej, P. A. *et al.* frazzled encodes a *Drosophila* member of the DCC immunoglobulin subfamily and is required for CNS and motor axon guidance. *Cell* **87**, 197–204 (1996).
- Su, M. *et al.* Regulation of the UNC-5 netrin receptor initiates the first reorientation of migrating distal tip cells in *Caenorhabditis elegans*. *Development* **127**, 585–594 (2000).
- Fox, R. M. *et al.* A gene expression fingerprint of *C. elegans* embryonic motor neurons. *BMC Genomics* **6**, 42 (2005).
- Sym, M., Robinson, N. & Kenyon, C. MIG-13 positions migrating cells along the anteroposterior body axis of *C. elegans*. *Cell* **98**, 25–36 (1999).
- Jin, Y., Jorgensen, E., Hartwig, E. & Horvitz, H. R. The *Caenorhabditis elegans* gene *unc-25* encodes glutamic acid decarboxylase and is required for synaptic transmission but not synaptic development. *J. Neurosci.* **19**, 539–548 (1999).
- Colavita, A., Krishna, S., Zheng, H., Padgett, R. W. & Culotti, J. G. Pioneer axon guidance by UNC-129, a *C. elegans* TGF- $\beta$ . *Science (New York, N. Y.)* **281**, 706–709 (1998).
- Sieburth, D. *et al.* Systematic analysis of genes required for synapse structure and function. *Nature* **436**, 510–517 (2005).
- Paganoni, S. & Ferreira, A. Expression and subcellular localization of Ror tyrosine kinase receptors are developmentally regulated in cultured hippocampal neurons. *J. Neurosci. Res.* **73**, 429–440 (2003).
- Phelan, P. Innexins: members of an evolutionarily conserved family of gap-junction proteins. *Biochim. Biophys. Acta* **1711**, 225–245 (2005).
- Whangbo, J. & Kenyon, C. A Wnt signaling system that specifies two patterns of cell migration in *C. elegans*. *Mol. Cell* **4**, 851–858 (1999).
- Colon-Ramos, D. A., Margeta, M. A. & Shen, K. Glia promote local synaptogenesis through UNC-6 (netrin) signaling in *C. elegans*. *Science* **318**, 103–106 (2007).
- Adler, C. E., Fetter, R. D. & Bargmann, C. I. UNC-6/Netrin induces neuronal asymmetry and defines the site of axon formation. *Nature Neurosci.* **9**, 511–518 (2006).
- Manitt, C., Wang, D., Kennedy, T. E. & Howland, D. R. Positioned to inhibit: netrin-1 and netrin receptor expression after spinal cord injury. *J. Neurosci. Res.* **84**, 1808–1820 (2006).
- McAllister, A. K. Dynamic aspects of CNS synapse formation. *Annu. Rev. Neurosci.* **30**, 425–450 (2007).
- Hall, A. C., Lucas, F. R. & Salinas, P. C. Axonal remodeling and synaptic differentiation in the cerebellum is regulated by WNT-7a signaling. *Cell* **100**, 525–535 (2000).
- Packard, M. *et al.* The *Drosophila* Wnt, wingless, provides an essential signal for pre- and postsynaptic differentiation. *Cell* **111**, 319–330 (2002).
- Umemori, H., Linhoff, M. W., Ornitz, D. M. & Sanes, J. R. FGF22 and its close relatives are presynaptic organizing molecules in the mammalian brain. *Cell* **118**, 257–270 (2004).
- McCabe, B. D. *et al.* The BMP homolog Gbb provides a retrograde signal that regulates synaptic growth at the *Drosophila* neuromuscular junction. *Neuron* **39**, 241–254 (2003).
- Herman, M. A., Vassilieva, L. L., Horvitz, H. R., Shaw, J. E. & Herman, R. K. The *C. elegans* gene *lin-44*, which controls the polarity of certain asymmetric cell divisions, encodes a Wnt protein and acts cell nonautonomously. *Cell* **83**, 101–110 (1995).
- Mello, C. & Fire, A. DNA transformation. *Methods Cell Biol.* **48**, 451–482 (1995).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the W. M. Keck Foundation, the McKnight Endowment Fund, the Searle Scholar Award and the Howard Hughes Medical Institute. We thank the International Caenorhabditis Genetic Center and M. Chalfie for strains. We further thank M. Chalfie for sharing unpublished results on the silencing system. We also thank C. Gao and Y. Fu for technical assistance, and T. Clandinin, C. Bargmann, S. McConnell, and members of the Shen laboratory for comments on the manuscript.

**Author Contributions** V.Y.P., M.P.K. and K.S. designed the experiments; V.Y.P. performed the experiments and data analysis; V.Y.P., M.P.K. and K.S. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to K.S. ([kangshen@stanford.edu](mailto:kangshen@stanford.edu)).



## METHODS

**Temperature shift experiments.** Animals were either cultured at 16 °C or 25 °C for multiple generations before being shifted to a different temperature. Experimental animals in the L3 and early L4 larval stages were placed at 16 °C for three days or 25 °C for two days before the phenotype was analysed. Scoring was performed in gravid adults with normal DA9 guidance for all experiments. *unc-5*; *mec-8* double mutants were analysed for experiments in Fig. 2d. It was later discovered that the temperature-dependent regulation of Ex[P*unc-5::intron::unc-5*] was independent of the *mec-8* mutation (Supplementary Fig. 8a).

**Constructs and transgenic worms.** *wyIs109*: a *XmaI*–*NheI* PCR fragment containing *cfp* was subcloned into *Pttx-3::rab-3* pSM obtained from *Pttx-3::mcherry::rab-3* (ref. 20) to make *Pttx-3::cfp::rab-3*. A *SphI*–*AsdI* fragment containing *Pmig-13* (ref. 3) was subcloned into *cfp::rab-3* pSM derived from *Pttx-3::cfp::rab-3* described above to make *Pmig-13::cfp::rab-3*. A *KpnI*–*Apal* fragment containing mCherry obtained from *Pmig-13::lin-17::mcherry*<sup>3</sup> was subcloned into *Pmig-13::gateway* pSM from *Pmig-13::gateway::yfp*<sup>3</sup> to make *Pmig-13::gateway::mcherry*. The *sng-1* entry clone was obtained from the ORFeome project (<http://worldb.dfci.harvard.edu/>) and cloned into the destination vector *Pmig-13::gateway::mcherry* using the gateway strategy with LR clonase (Invitrogen) to make *Pmig-13::sng-1::mcherry*. *Pmig-13::cfp::rab-3*, *Pmig-13::sng-1::mcherry* and *Pmig-13::snb-1::yfp*<sup>3</sup> were injected with *Podr-1::gfp* at 20 ng  $\mu\text{l}^{-1}$  and integrated into chromosome V using trimethylpsoralen/ultraviolet mutagenesis. *cfp* primers: 5'-TCCCCCGGGATGAGTAAAGGAGAAGAAGCTTTTCAC and 3'-CTAGCTAGCTTTGTATAGTTCATCCATGCCATG.

*wyEx2055*: a *NheI*–*KpnI* fragment containing the *syd-2* genomic sequence from *Pgcy-8::mCherry::syd-2* was subcloned into *Pitr-1 pB::gfp* pSM to make *Pitr-1 pB::gfp::syd-2*; a *FseI*–*AsdI* PCR fragment containing *Pitr-1 pB* was subcloned into *mcherry::rab-3* pSM from *Pglr-3::mcherry::rab-3*. *Pitr-1 pB::gfp::syd-2* was injected at 0.5 ng  $\mu\text{l}^{-1}$  with *Pitr-1 pB::mcherry::rab-3* at 10 ng  $\mu\text{l}^{-1}$  with *Podr-1::gfp* at 20 ng  $\mu\text{l}^{-1}$  into N2 animals. *Pitr-1 pB* primers: 5'-GAAAGGGGCGCCATCTATTTCCAGAGTTCGTTCCCGAGC and 3'-CTTTCGGCGCGCGCCAAATTCGTGTGCTTCCACCACCAC.

*wyEx1902*: a *SphI*–*AsdI* PCR fragment containing *Pitr-1 pB* was subcloned into *mcherry::gateway* from *Pmig-13::mcherry::gateway*. *Pitr-1 pB::mcherry::gateway* was injected at 5 ng  $\mu\text{l}^{-1}$  with *Podr-1::gfp* at 20 ng  $\mu\text{l}^{-1}$  into N2 animals. *Pitr-1 pB* primers: 5'- (ref. 3) and 3'-GAAAGGGGCGCGCCCAATTCGTGTGCTTCACCAC.

*wyEx1311*, *wyEx1485*: a *KpnI*–*Apal* fragment containing the *unc-5* 3' untranslated region (UTR) was subcloned into *Pmig-13::snb-1* pSM and a *AsdI*–*KpnI* PCR fragment containing the *unc-5* genomic sequence was subcloned into *Pmig-13::unc-5* 3' UTR pSM. The *Pmig-13::unc-5* plasmid was injected at 2 ng  $\mu\text{l}^{-1}$  with *Podr-1::gfp* at 20 ng  $\mu\text{l}^{-1}$  into *unc-5*; *wyIs85* mutants. The two arrays are separate lines obtained in independent injections. Higher levels of the plasmid were toxic and there would be no germline transmission, whereas lower levels did not rescue the defect. *unc-5* primers: 5'-GAAAGGGGCGCGCCCATGACGAAATCACAATCACACACAAC and 3'-GAAGGGGTACCAGTGGGGACACAATTTGTGGAAGAGCTG; *unc-5* 3' UTR primers: 5'-GAAAGGGGTACCGTCAATTTTGTGCACAAACACAACACTAG and 3'-GAAAGGGGGCCCGGTCTTTCTGCATAGAAAATCGC.

*wyEx1498*: an *AsdI*–*KpnI* PCR fragment containing the *unc-5* genomic sequence and a *SphI*–*AsdI* PCR fragment containing the *Punc-5* were subcloned

into *sl2 dsred* pSM<sup>20</sup> to make *Punc-5::unc-5::sl2::dsred*. This plasmid was injected at 10 ng  $\mu\text{l}^{-1}$  with *Podr-1::gfp* at 20 ng  $\mu\text{l}^{-1}$  into *unc-5*; *lin-18*; *wyIs85* mutants.

*wyEx1228*, *wyEx2418*, *wyEx2419*: a *SphI*–*AsdI* PCR fragment containing *Punc-5* was subcloned into *unc-5::unc-5* 3' UTR from *Pmig-13::unc-5::unc-5* 3' UTR. *Punc-5::unc-5* was injected at 10 ng  $\mu\text{l}^{-1}$  with *Podr-1::gfp* at 20 ng  $\mu\text{l}^{-1}$  into *unc-5*; *wyIs85* mutants (*wyEx1228*). *Punc-5::unc-5* was injected at 8 ng  $\mu\text{l}^{-1}$  with *Pmig-13::mcherry::gateway*<sup>3</sup> at 8 ng  $\mu\text{l}^{-1}$  with *Podr-1::gfp* at 20 ng  $\mu\text{l}^{-1}$  into *unc-5*; *wyIs85* mutants. *wyEx2418* and *wyEx2419* were two arrays obtained that had animals where UNC-5 was absent in DA9. Most of the other lines obtained had no animals where UNC-5 was absent in DA9. *Punc-5* primers: 5'-GAAAGGGCATGCTGAGCTTTTCCAACTAGAGAGCTTC and 3'-GAAAGGGGCGCCTACTGGAATAGAAATTATGATTAGTGACAACTTG.

*wyEx1277*: an *EcoRI*–*Apal* PCR fragment containing the *unc-5* 3' UTR was subcloned into *Pmig-13::snb-1::yfp* pSM<sup>3</sup> to make *Pmig-13::snb-1::yfp::unc-5* 3' UTR. An *AsdI*–*KpnI* PCR fragment containing the *unc-5* genomic sequence (similar to *wyEx1311*) and a *SphI*–*AsdI* PCR fragment containing *Pitr-1 pB* (similar to *wyEx2055*) were subcloned into *unc-5* 3' UTR pSM derived from make *Pmig-13::snb-1::yfp::unc-5* 3' UTR. *Pitr-1 pB::unc-5::yfp* was injected at 80 ng  $\mu\text{l}^{-1}$  with *Podr-1::gfp* at 20 ng  $\mu\text{l}^{-1}$  into N2 animals. *unc-5* 3' UTR primers: 5'-GAAAGGGAATTCGCTCAATTTTTCACAAACACAACACTAG and 3'-GAAAGGGGGCCCGGTCTTTCTGCATAGAAAATCGC.

*wyEx1904*, *wyEx1916*: a *SphI*–*AsdI* PCR fragment containing *Pegl-20* was subcloned into *unc-6* pSM derived from *Punc-6::unc-6* (ref. 20). The *Pegl-20::unc-6* plasmid was injected at 20 ng  $\mu\text{l}^{-1}$  with *Pttx-3::cfp* at 50 ng  $\mu\text{l}^{-1}$  into N2 animals. The two arrays were separate lines obtained from one injection. *Pegl-20* primers: 5'-GAAAGGGCATGCAAGTTCCCTTTTATTTTGAAGTCA TCC and 3'-GAAAGGGGCGCGCCTATTCTGAAATTGAGATGTTT TAG AATTTC.

*wyEx2093*, *wyEx2094*: an *AsdI*–*KpnI* PCR fragment containing the *lin-44* complementary DNA was subcloned into *Punc-6* pSM derived from *Punc-6::unc-6::mcherry*<sup>20</sup>. The *Punc-6::lin-44* plasmid was injected at 20 ng  $\mu\text{l}^{-1}$  with *Podr-1::gfp* at 20 ng  $\mu\text{l}^{-1}$  into *unc-5*; *wyIs85* mutants. The two arrays were separate lines obtained in the same injection. *lin-44* primers: 5'-GAAAGGGGCGCGCCATGCGAGCAGCTCTTTTGATTTC and 3'-GAAAGGGGTACCTTAA AAAATTAGGCTTTTTCGGCGGTG.

*wyEx2306*, *wyEx2308*: an *AsdI* PCR fragment obtained from pAC13, a gift from M. Chalfie, containing *mec-2* *intron9* was subcloned into *Punc-5::unc-5* (similar to *wyEx1494*). This plasmid was injected at 10 ng  $\mu\text{l}^{-1}$  with *Podr-1::gfp* at 20 ng  $\mu\text{l}^{-1}$  into *unc-5*; *mec-8*; *wyIs85* mutants. The two arrays were separate lines obtained from one injection. *unc-5*; *wyIs85*; *wyEx2308* mutant animals were obtained by crossing *unc-5*; *mec-8*; *wyIs85*; *wyEx2308* animals with N2 animals. *mec-2* *intron9* primers: 5'-GAAAGGGGCGCGCCACCGCCTAAAGTG TAAGTTTTC and 3'-GAAAGGGGCGCGCCGACGGTGGCTCCTCACTGAA AAC.

*wyEx1054*, *wyEx2396*, *wyEx2430*: the *unc-9*, *F35D2.3* and *dys-1* entry clones were obtained from the ORFeome project (<http://worldb.dfci.harvard.edu/>) and cloned into the destination vector *Pitr-1 pB::gateway::yfp*<sup>3</sup> using the gateway strategy with LR clonase (Invitrogen) to make *Pitr-1 pB::unc-9::yfp*, *Pitr-1 pB::F35D2.3::yfp* and *Pitr-1 pB::dys-1::yfp*. These plasmids were then injected at 10 ng  $\mu\text{l}^{-1}$  or 40 ng  $\mu\text{l}^{-1}$  (for *dys-1*) with *Podr-1::dsRED* or *Podr-1::GFP* at 20 ng  $\mu\text{l}^{-1}$  into N2 worms.

## LETTERS

# STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling

Hiroki Ishikawa<sup>1</sup> & Glen N. Barber<sup>1</sup>

The cellular innate immune system is essential for recognizing pathogen infection and for establishing effective host defence. But critical molecular determinants responsible for facilitating an appropriate immune response—following infection with DNA and RNA viruses, for example—remain to be identified. Here we report the identification, following expression cloning, of a molecule (STING; stimulator of interferon genes) that appears essential for effective innate immune signalling processes. It comprises five putative transmembrane regions, predominantly resides in the endoplasmic reticulum and is able to activate both NF- $\kappa$ B and IRF3 transcription pathways to induce expression of type I interferon (IFN- $\alpha$  and IFN- $\beta$ ) and exert a potent anti-viral state following expression. In contrast, loss of STING rendered murine embryonic fibroblasts extremely susceptible to negative-stranded virus infection, including vesicular stomatitis virus. Further, STING ablation abrogated the ability of intracellular B-form DNA, as well as members of the herpesvirus family, to induce IFN- $\beta$ , but did not significantly affect the Toll-like receptor (TLR) pathway. Yeast two-hybrid and co-immunoprecipitation studies indicated that STING interacts with RIG-I and with SSR2 (also known as TRAP $\beta$ ), which is a member of the translocon-associated protein (TRAP) complex required for protein translocation across the endoplasmic reticulum membrane following translation<sup>1,2</sup>. Ablation by RNA interference of both TRAP $\beta$  and translocon adaptor SEC61 $\beta$  was subsequently found to inhibit STING's ability to stimulate expression of IFN- $\beta$ . Thus, as well as identifying a regulator of innate immune signalling, our results imply a potential role for the translocon in innate signalling pathways activated by select viruses as well as intracellular DNA.

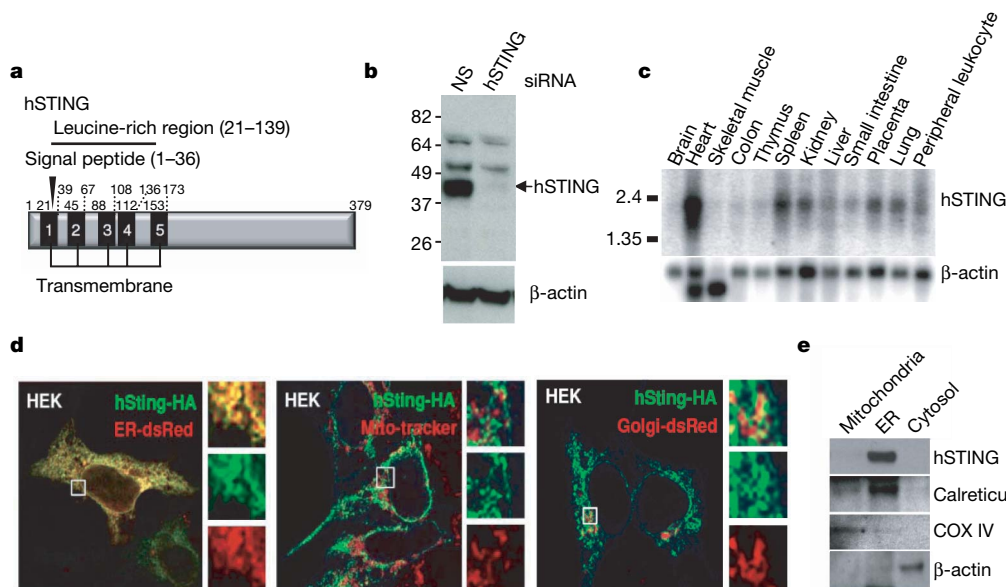
Cellular host defence responses to pathogen invasion principally involve the detection of pathogen associated molecular patterns (PAMPs), such as viral nucleic acid or bacterial cell wall components (including lipopolysaccharide or flagellar proteins), that results in the induction of anti-pathogen genes<sup>3–9</sup>. For example, viral RNA can be detected by membrane bound TLRs present in the endoplasmic reticulum (ER) and/or endosomes (for example, TLR3 and TLR7/8) or by TLR-independent intracellular DExD/H box RNA helicases, referred to as retinoic acid inducible gene 1 (RIG-I) or melanoma differentiation associated antigen 5 (MDA5, also referred to as IFIH1 and helicard)<sup>3–10</sup>. Pathogen DNA can be recognized by TLR9 present in plasmacytoid dendritic cells, although it is now apparent that important TLR-independent pathways also exist to recognize DNA in alternative tissue, the mechanisms of action of which remain to be determined<sup>3–10</sup>. These events culminate in the activation of downstream signalling events, leading to the transcription of NF- $\kappa$ B and IRF3/7-dependent genes, including type I IFN.

To further determine the mechanisms of innate immune signalling, we employed an expression screening system in which approximately

5,500 human and 9,000 murine full length complementary DNAs were individually transfected into 293T cells harbouring a luciferase gene under control of the IFN- $\beta$  promoter (IFN- $\beta$ -Luc). The top five hits whose overexpression lead to the significant induction of IFN- $\beta$ -Luc were found to be IPS-1 (also referred to as VISA/CARDIF/MAVS) (Supplementary Fig. 1)<sup>11–14</sup>. However, we also isolated a previously uncharacterized molecule (gi:38093659/NP\_938023/2610307O08RIK) which we referred to as STING that harboured five predicted transmembrane motifs (in humans) and existed as a 379 amino acid protein in human cells and 378 amino acids in murine cells (Fig. 1a and Supplementary Fig. 1). A putative signal cleavage motif was found to exist at position 1–36 and a leucine rich region was apparent between amino acids 21 and 139 (Fig. 1a). The predicted molecular weight of human STING was 42,192 Da, which approximately corresponded to its observed molecular weight in human 293 cells following immunoblot analysis using a rabbit antiserum raised to a STING peptide (Fig. 1b). RNAi studies confirmed that the observed 42 kDa band was indeed STING (Fig. 1b). STING was found to be ubiquitously expressed in a variety of tissues, as determined by northern analysis, and was found to predominantly reside in the ER region of the cell as determined by confocal microscopy and fractionation analysis (Fig. 1c–e).

Overexpression of STING in 293T cells was subsequently confirmed to robustly induce the expression of the IFN promoter (IFN- $\beta$ -Luc) by up to 400-fold, but not a control TK promoter driving luciferase (pRL-TK), interferon regulatory factor 3 (IRF3) responsive promoters (PRD-III-I-Luc) up to 1,000-fold, an NF- $\kappa$ B responsive promoter (NF- $\kappa$ B-Luc) 12-fold, and interferon-inducible promoters (interferon sensitive response element, ISRE-Luc) up to 800-fold (Fig. 2a–d). STING did not activate control promoters driving luciferase reporters such as those derived from the RB, p53 or E2F genes (Supplementary Fig. 2). Increased dimerization of IRF3 was also observed in STING expressing 293T cells, confirming that STING regulates the induction of type I IFN at locations upstream of IRF3 activation, dimerization and translocation (Supplementary Fig. 2)<sup>15</sup>. Endogenous *Ifnb* mRNA and IFN- $\beta$  protein was induced significantly in murine embryonic fibroblasts (MEFs) transiently transfected with STING (Fig. 2e, f). DNA microarray analysis of STING expression in 293T cells further emphasized STING's ability to induce primary innate immune response genes (Fig. 2g). Accordingly, MEF cells expressing STING or IPS-1 were significantly resistant to vesicular stomatitis virus (VSV) infection (Fig. 2h, i). Subsequent analysis indicated that STING function was ablated in the absence of the I $\kappa$ B kinase family member TBK-1, confirming that STING's activity involved activation of IRF3 and was indeed upstream of this kinase (Fig. 2j)<sup>15</sup>. Finally, we observed that STING did not exert robust activity in the absence of FADD, which has also been shown to be important for efficient innate immune signalling processes (Supplementary Fig. 2)<sup>16</sup>.

<sup>1</sup>Department of Medicine and Sylvester Comprehensive Cancer Center, University of Miami School of Medicine, Miami, Florida 33136, USA.



**Figure 1 | STING is an ER protein.** **a**, Schematic of human STING (hSTING) indicating transmembrane and leucine rich regions. **b**, Immunoblot analysis of STING in HEK 293 cells treated with RNAi to STING (hSTING) or control RNAi (NS). **c**, Northern blot analysis of human STING and control β-actin. **d**, Confocal analysis of HEK 293 cells (HEK) transfected with human

STING tagged at the carboxyl end with HA. Transfected cells were also analysed using ER-dsRed, Mito Tracker Red or Golgi-dsRed. **e**, Fractionation experiments confirm that STING resides in the ER. Control antibodies indicate accuracy of fractionation (Calreticulin, ER; COX IV, mitochondria; β-actin, cytosol).

We also established that while STING that had been tagged with HA at the carboxy region retained activity, this activity was lost when STING was tagged at the amino terminus or C terminus with GFP (data not shown). Subsequent analysis indicated that the N-terminal region of STING containing the five putative transmembrane regions (amino acids 1–230) or just the carboxyl region of STING (amino acids 173–379) did not exhibit significant ability, alone, to induce the IFN-β-Luc promoter (Fig. 2k, l). Thus, full length, intact STING is required for efficient function. However, we further observed that the carboxyl region of STING exerted a dominant-negative inhibitory effect and could impede the ability of full-length STING to stimulate IFN-β-Luc (Fig. 2m). Collectively, these data indicate that expression of STING activates the innate immune response, including type I IFN, leading to the induction of an antiviral state.

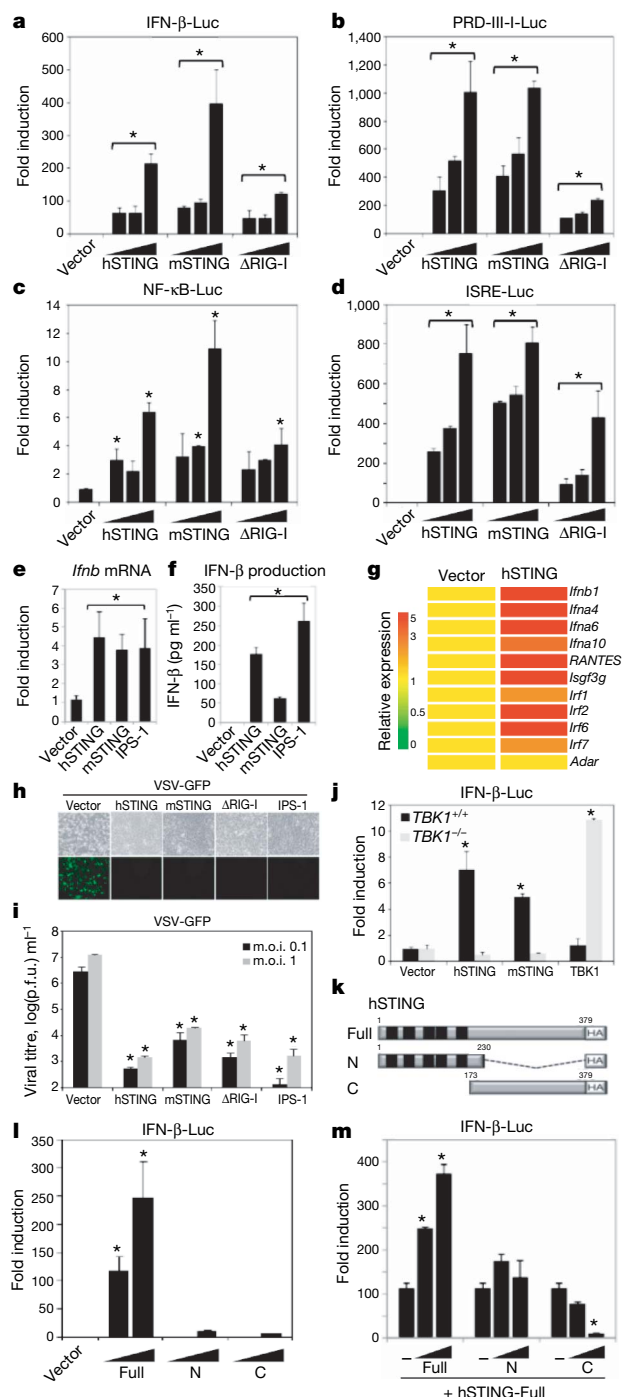
To further analyse STING function, we used an RNAi approach to ablate STING in a number of cell types. Our data indicated that knockdown of STING in HEK 293 cells modestly reduced the ability of the negative-stranded rhabdovirus VSV-GFP to induce IFN-β, presumably because these viruses are only weak activators of IFN-β (Supplementary Fig. 3). However, such cells were rendered extremely susceptible to virus infection and replication. To confirm a requirement for STING in the regulation of type I IFN induction and in host defence, we generated STING deficient (*Sting*<sup>−/−</sup>) mice by targeted homologous recombination in embryonic stem (ES) cells (Supplementary Fig. 4). *Sting*<sup>−/−</sup> animals were born at the Mendelian ratio, and developed and bred normally. Accordingly, MEFs from wild-type and *Sting*<sup>−/−</sup> animals were infected with VSV-GFP at varying multiplicity of infection (m.o.i., 0.01–1) for up to 36 h post infection. This study confirmed that more progeny virus was produced in MEFs lacking STING compared to controls (by a factor of 10<sup>2</sup>; 24–36 h) (Fig. 3a–d). These data were verified using VSV expressing a luciferase reporter gene and VSV-ΔM, which exhibits a defect in the viral matrix protein normally responsible for inhibiting cellular messenger RNA export from the nucleus<sup>17</sup> (Supplementary Fig. 4 and Fig. 3e). Reconstitution of STING to *Sting*<sup>−/−</sup> MEFs rescued the susceptibility to VSV infection (Supplementary Fig. 4). Similar analysis also indicated that loss of STING also reduced the ability of Sendai virus to induce IFN-β

(Fig. 3f). In contrast, we did not observe a strong requirement for STING to mediate the ability of transfected poly I:C to induce IFN-β induction, which is largely governed by the intracellular RIG-I homologue MDA5 (ref. 18) (Supplementary Fig. 5). We also observed that the positive-stranded encephalomyocarditis virus (EMCV), a member of the picornavirus family, did not effectively induce IFN-β or replicate differently, regardless of the presence of STING (Supplementary Fig. 5).

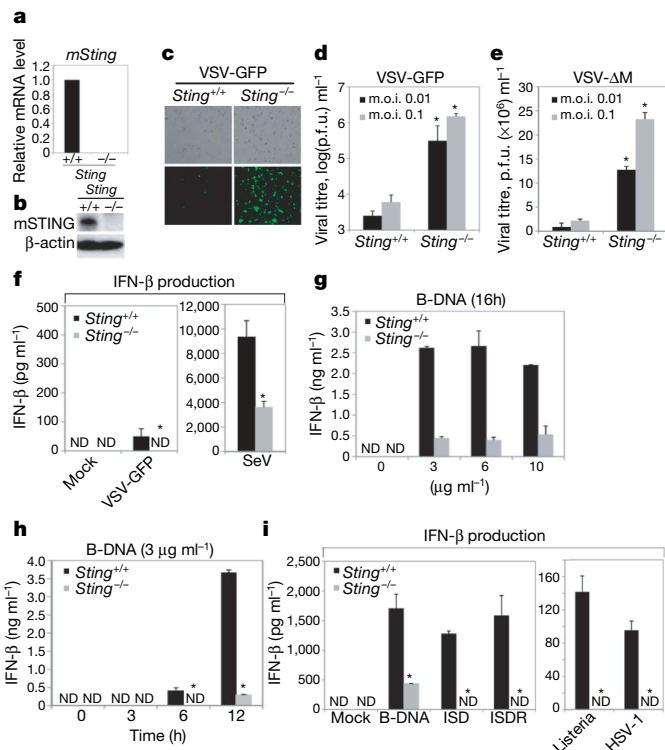
Thus we conclude that STING may play a more predominant role in facilitating RIG-I mediated innate signalling, rather than MDA5. Interestingly, we did notice a significant defect (>5-fold) in the ability of transfected B-form DNA (poly dA-dT) to induce IFN-β in MEFs lacking STING compared to controls (Fig. 3g, h). More strikingly, the non CpG containing interferon stimulatory DNA (ISD) was completely unable to induce IFN-β in *Sting*<sup>−/−</sup> MEFs, as were the DNA virus herpes simplex virus 1 and bacteria *Listeria monocytogenes* (Fig. 3i). TLR9 is considered to govern CpG DNA-mediated induction of IFN-β, but is not active in MEFs<sup>19,20</sup>. Thus, it is plausible that STING may function in TLR9-independent, DNA-mediated induction of type I IFN. This effect was similarly observed in murine STING-lacking bone marrow derived macrophages (BMDM) or bone marrow derived dendritic cells (GM-DC) cultured using granulocyte-macrophage colony stimulating factor (GM-CSF) (Supplementary Fig. 6). However, no significant difference was observed in the ability of exogenous poly I:C or lipopolysaccharide (LPS) to induce IFN-β, when comparing *Sting*<sup>−/−</sup> BMDMs or GM-DCs to controls, events which depend on TLR3 and TLR4, respectively. Whereas loss of STING rendered MEFs highly susceptible to VSV-GFP, less susceptibility was observed following VSV-GFP infection of *Sting*<sup>−/−</sup> GM-DCs or BMDMs, indicating that STING may be more important in facilitating negative-stranded virus-mediated innate signalling in fibroblasts. Collectively, our data indicate that loss of STING leads to a defect in RIG-I mediated type I IFN induction but does not affect the TLR pathway. In addition, we report that STING functions in the pathway used by intracellular B-form DNA to induce IFN-β.

To further examine the mechanisms of STING's function in innate signalling, we attempted to determine if STING interacted with RIG-I



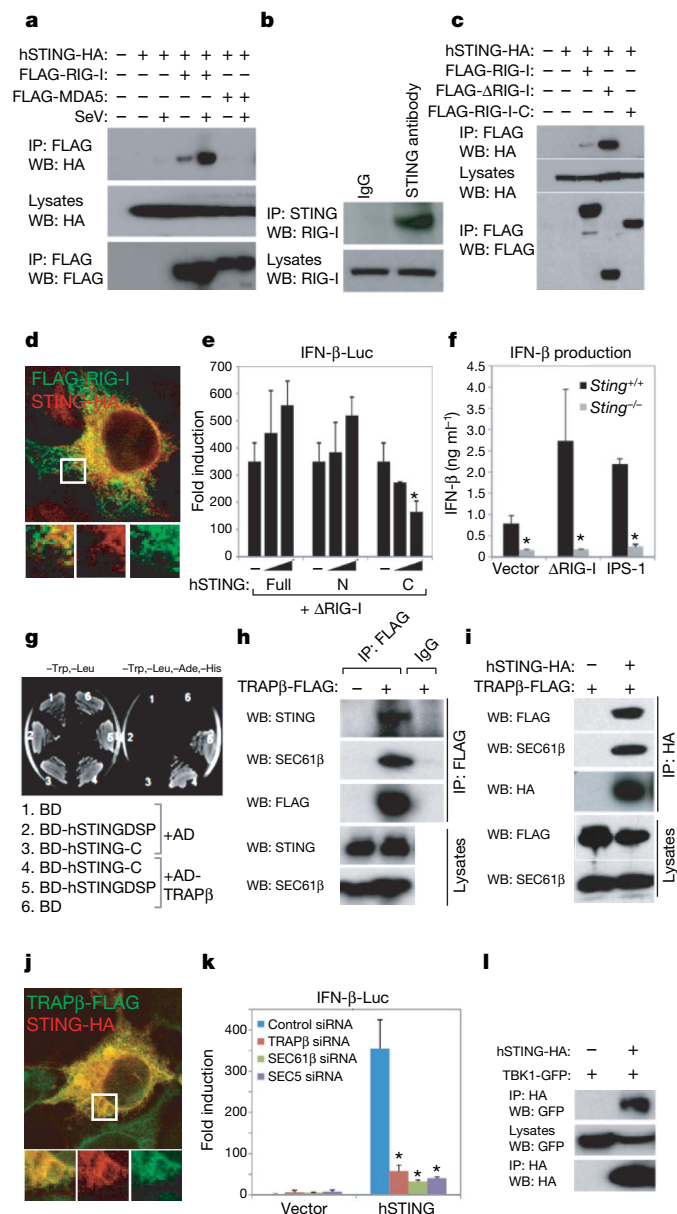


**Figure 2 | STING facilitates IFN induction.** **a**, 293T cells were transfected with an IFN- $\beta$ -Luc (p110-Luc) plasmid and increasing amounts of human STING (hSTING), murine STING (mSTING) or control  $\Delta$ RIG-I. **b–d**, 293T cells transfected as in **a** with either PRD-III-I-Luc (**b**), NF- $\kappa$ B-Luc (**c**) or ISRE-Luc (**d**) reporter plasmids were analysed similarly. **e**, MEFs were transfected as in **a**, and *Irfn* mRNA was analysed by qRT-PCR. **f**, Medium from transfected MEFs was analysed for IFN- $\beta$  protein by ELISA. **g**, Microarray analysis of 293T cells transfected with hSTING. **h**, MEFs transfected with STING or  $\Delta$ RIG-I or IPS-1 are resistant to VSV-GFP infection (m.o.i., 1). Top panel is light microscopy, lower panel is fluorescent microscopy. **i**, Plaque assay from **h**. p.f.u., plaque-forming units. **j**, TBK-1 deficient MEFs do not facilitate STING signalling. **k**, Schematic of hSTING variants. **l**, 293T cells were transfected as in **a** with hSTING or variants, and luciferase measured. **m**, 293T cells were transfected with STING and increasing amounts of full length STING (hSTING-Full), hSTING-N or hSTING-C with luciferase plasmids as in **a**. Asterisks indicate significant difference ( $P < 0.05$ ) as determined by Student's *t*-test. Error bars indicate s.d.



**Figure 3 | Loss of STING affects host defence.** **a**, qRT-PCR analysis of murine *Sting* mRNA in *Sting*<sup>-/-</sup> or control MEFs. **b**, Immunoblot of murine STING in -/- cells or controls. **c**, Fluorescence microscopy (GFP) of *Sting*<sup>-/-</sup> or controls infected with VSV-GFP (m.o.i., 0.1). **d**, Viral titres from **c**. **e**, Viral titres following VSV- $\Delta$ M infection. **f**, Endogenous IFN- $\beta$  levels from *Sting*<sup>-/-</sup> or controls after infection with VSV-GFP (m.o.i., 1) or Sendai virus (SeV; m.o.i., 1). **g**, *Sting*<sup>-/-</sup> MEFs or controls were treated with transfected B-DNA and IFN- $\beta$  measured by ELISA. **h**, Time-course analysis of **g**. **i**, *Sting*<sup>-/-</sup> or controls were exposed to transfected B-form DNA, interferon stimulatory DNA (ISD), ISD reversed nucleotide sequence (ISDR), *Listeria monocytogenes* (m.o.i. 50) and HSV-1 (m.o.i. 5) for 12 h, and IFN- $\beta$  measured by ELISA. Asterisks indicate significant difference ( $P < 0.05$ ) as determined by Student's *t*-test. Error bars indicate s.d. ND, not detectable.

and MDA5, which are putative innate immune signalling receptors for negative- or positive-stranded viral RNA, respectively<sup>3,4,10</sup>. Co-immunoprecipitation experiments in 293T cells principally indicated that FLAG-tagged RIG-I but not MDA5 could associate with HA-tagged STING in co-transfection experiments (Fig. 4a). The binding of RIG-I to STING was augmented upon infection of 293T cells with Sendai virus (Fig. 4a). We confirmed in normal human umbilical vein endothelial cells (HUVECs) that endogenous RIG-I could associate directly, or indirectly as a complex, with endogenous STING (Fig. 4b). This is in agreement with previous data indicating that STING seems to preferentially modulate the RIG-I, rather than the MDA5, regulated pathway (Fig. 3d and Supplementary Fig. 5). We also determined that the CARD domains of RIG-I (amino acids 1–284) were preferentially able to associate with HA-STING in transfected 293T cells (Fig. 4c). RIG-I was subsequently shown to colocalize with STING in the co-transfected 293T cells (Fig. 4d). It was similarly observed that the carboxyl region of STING could inhibit the function of RIG-I, again indicating that this region of STING can exhibit a dominant-inhibitory effect (Fig. 4e). We also noted association of the RIG-I downstream adaptor IPS-1 with STING although, similarly to the situation with RIG-I, it is not yet clear whether IPS-1 directly interacts with STING or exists as a complex with RIG-I/STING (Supplementary Fig. 7). Accordingly, we observed that STING was able to induce the expression of an IFN- $\beta$  driven luciferase construct in MEFs lacking RIG-I or IPS-1, probably confirming that that STING functions downstream of these latter molecules



**Figure 4 | STING associates with the translocon.** **a**, 293T cells were co-transfected with HA-STING, FLAG-RIG-I or MDA5 and infected with Sendai virus (SeV; m.o.i. 1). Lysates were immunoprecipitated (IP) and immunoblotted (IB) using antibodies to HA or FLAG. **b**, Endogenous human STING associates with RIG-I in HUVECs. **c**, ΔRIG-I (amino acids 1–284) and not RIG-I-C (amino acids 218–925) associate with STING in co-transfected 293T cells. **d**, Confocal image of 293T cells co-transfected with tagged STING and RIG-I. **e**, 293T cells were co-transfected with control vector (–) or increasing amounts of full-length, amino (amino acids 1–230) or carboxyl (amino acids 173–379) STING and ΔRIG-I, and IFN-β-Luc was measured. **f**, Control or *Sting*<sup>–/–</sup> MEFs were transfected with ΔRIG-I (amino acids 1–284) or IPS-1 and IFN-β was measured by ELISA. **g**, GAL4 binding domain (BD) fused to the carboxyl region of hSTING, BD-hSTING-C, interacts with Ssr2/TRAPβ fused to the GAL4 activation domain (AD-hTRAPβ) in yeast-two hybrid screening (BD-hSTINGASP, amino acids 36–369; BD-hSTING-C, amino acids 173–379). **h**, HEK 293 cells were transfected with FLAG-tagged TRAPβ and endogenous STING or SEC61β measured by immunoblot. **i**, STING and TRAPβ were co-transfected HEK 293 cells and analysis carried out as in **h**. **j**, Co-localization of STING and TRAPβ in 293T cells. **k**, RNAi to TRAPβ, SEC61β or SEC5 in HEK 293 cells ablates STING signalling. **l**, HA-STING associates with GFP-TBK1 in co-transfected HEK 293 cells. Asterisks indicate significant difference ( $P < 0.05$ ) as determined by Student's *t*-test. Error bars indicate s.d.

(Supplementary Fig. 8). The ability of ΔRIG-I or IPS-1 to induce IFN-β appeared diminished in *Sting*<sup>–/–</sup> MEFs (Fig. 4f). However, there was also a marked reduction in the induction of IFN-β by all transfected plasmids (including vector alone) in the absence of STING compared to control MEFs, probably because endogenous DNA innate signalling pathways are defective (Fig. 3g–i). These data indicate that STING may be an important downstream adaptor molecule that facilitates RIG-I and perhaps IPS-1 function.

To gain further insight into the molecular mechanisms of the action of STING, we screened an IFN-induced, human fibroblast yeast two-hybrid cDNA library using STING (amino acids 173–379) as a bait and repeatedly isolated SSR2/TRAPβ, a member of the TRAP complex comprising four subunits (α–Δ) that facilitates translocation of proteins into the ER following translation (Fig. 4g)<sup>1,2</sup>. The TRAP complex is known to associate with the translocon, comprising three subunits, SEC61α, SEC61β and SEC61γ. Given this information, we confirmed that TRAPβ can indeed associate with endogenous STING in HEK 293 cells following co-immunoprecipitation experiments (Fig. 4h). Using this approach, we confirmed that TRAPβ also co-immunoprecipitated with endogenous SEC61β (Fig. 4h). We next verified that STING could also associate not only with TRAPβ but also SEC61β, probably as a complex (Fig. 4i). STING was also observed to colocalize with TRAPβ in the ER region of the cell (Fig. 4j). Indeed, loss of TRAPβ or SEC61β reduced STING's ability to induce an IFN-β promoter driving luciferase (Fig. 4k; Supplementary Fig. 9). Taken together, these data indicate that the STING may be involved in translocon function, and that the translocon may be able to influence the induction of type I IFN.

Thus, STING is predominantly an ER resident protein that may link RIG-I and DNA-mediated intracellular innate signalling to the translocon. We speculate that RIG-I may detect translating viral RNAs at the intersection of ribosome/ER translocon association and require STING to exert effective function<sup>2</sup>. Alternatively, STING may participate in mediating ER stress response pathways, but this remains to be verified. Although it is not clear how signalling from the translocon to IRF3/NF-κB occurs, it has recently been established that the translocon physically associates with the exocyst—the octameric Sec6–Sec8 complex that also associates with the ER and tethers secretory vesicles to membranes, and facilitates protein synthesis and secretion<sup>21,22</sup>. Recently, the exocyst complex was found to recruit and activate TBK1 and play a role in type I IFN-β induction<sup>23</sup>. Our preliminary analysis indicates that STING also co-immunoprecipitates with TBK1, and that RNAi ablation of Sec5 also rendered cells defective in the STING function (Fig. 4k, l). Thus STING may facilitate the detection of intracellular viral RNA species as well as B-form DNA, indicating convergence of these intracellular PAMP recognition pathways.

## METHODS SUMMARY

**Plasmid constructs.** Human STING (hSTING), murine STING (mSTING), hSTING-N (amino acids 1–230) and hSTING-C (amino acids 173–379) sequences were amplified by PCR and were cloned into pcDNA3 plasmids to generate C-terminally HA-tagged expression constructs. Expression plasmids encoding Flag-tagged RIG-I, ΔRIG-I (amino acids 1–284), IPS-1 and TBK1 were described previously<sup>24</sup>. GFP-tagged RIG-I and TBK1 were generated by cloning into pAcGFP1-C1 (Clontech). Other plasmids were obtained as follows: p110-Luc and PRD-III-I-Luc (T. Maniatis), IFN-β-Luc and ISRE-Luc (J. Hiscott), NFκB-Luc (Stratagene), ER-dsRED and Golgi-dsRED (Clontech).

**Antibodies.** Rabbit polyclonal antibody against a synthetic peptide corresponding to residues 324–340 of human STING was obtained from EvoQuest Custom Antibody Services (Invitrogen). Other antibodies were obtained from following sources: SEC61β (Upstate), β-Actin, HA, FLAG (Sigma), COX IV, Calreticulin, IRF3 and RIG-I (Abcam).

**Confocal microscopy.** ER-dsRED and Golgi-dsRED (Clontech) were used for ER and Golgi marker, respectively. For mitochondria staining, living cells were incubated with 300 nM of Mito Tracker Red (Invitrogen) for 45 min at 37 °C.

**RNA interference.** Chemically synthesized 21-nucleotide siRNA duplexes were obtained from Dharmacon. RNA oligonucleotides used for human and murine STING were as follows: hSTING, GCAUCAAGGAUCGGGUUU; mSTING, CCAACAGCGUCUACGAGA.

**Generation of the *Sting* knockout mice.** The linearized targeting vector was electroporated into E14.1 ES cells originated from 129SvEv strain, followed by the selection in G418. One positive clone was injected into C57BL/6J blastocysts and *Sting*<sup>-/-</sup> mice generated on a C57BL/6J background.

**Yeast two-hybrid analyses.** To screen for interacting partners of STING, a yeast two-hybrid approach was used, adopting a novel IFN-induced, telomerase immortalized human fibroblast cDNA library, generated by our laboratory. hSTING-C (amino acids 173–379) sequence was amplified by PCR and was cloned into pGBK-T7 (Clontech) to generate a bait plasmid pGBK-T7-hSTING-C. 100 µg of the IFN-induced library's plasmid DNA was used to perform a library-scale transformation of yeast strain Y187.

**Statistics.** Students *t*-test was used to analyse data.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 7 July; accepted 6 August 2008.**

**Published online 24 August 2008.**

- Hartmann, E. *et al.* A tetrameric complex of membrane proteins in the endoplasmic reticulum. *Eur. J. Biochem.* **214**, 375–381 (1993).
- Menetret, J. F. *et al.* Architecture of the ribosome-channel complex derived from native membranes. *J. Mol. Biol.* **348**, 445–457 (2005).
- Takeuchi, O. & Akira, S. Recognition of viruses by innate immunity. *Immunol. Rev.* **220**, 214–224 (2007).
- Beutler, B. *et al.* Genetic analysis of resistance to viral infection. *Nature Rev. Immunol.* **7**, 753–766 (2007).
- Takahashi, K. *et al.* Nonspecific RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. *Mol. Cell* **29**, 428–440 (2008).
- Pichlmair, A. *et al.* RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* **314**, 997–1001 (2006).
- Hornung, V. *et al.* 5'-Triphosphate RNA is the ligand for RIG-I. *Science* **314**, 994–997 (2006).
- Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5**, 730–737 (2004).
- Loo, Y. M. *et al.* Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *J. Virol.* **82**, 335–345 (2008).
- Onomoto, K., Yoneyama, M. & Fujita, T. Regulation of antiviral innate immune responses by RIG-I family of RNA helicases. *Curr. Top. Microbiol. Immunol.* **316**, 193–205 (2007).
- Kawai, T. *et al.* IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nature Immunol.* **6**, 981–988 (2005).
- Meylan, E. *et al.* Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167–1172 (2005).
- Seth, R. B. *et al.* Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF- $\kappa$ B and IRF 3. *Cell* **122**, 669–682 (2005).
- Xu, L. G. *et al.* VISA is an adapter protein required for virus-triggered IFN- $\beta$  signaling. *Mol. Cell* **19**, 727–740 (2005).
- McWhirter, S. M. *et al.* IFN-regulatory factor 3-dependent gene expression is defective in Tbk1-deficient mouse embryonic fibroblasts. *Proc. Natl Acad. Sci. USA* **101**, 233–238 (2004).
- Balachandran, S., Thomas, E. & Barber, G. N. A FADD-dependent innate immune mechanism in mammalian cells. *Nature* **432**, 401–405 (2004).
- Faria, P. A. *et al.* VSV disrupts the Rael1/mrnp41 mRNA nuclear export pathway. *Mol. Cell* **17**, 93–102 (2005).
- Kato, H. *et al.* Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441**, 101–105 (2006).
- Ishii, K. J. *et al.* A Toll-like receptor-independent antiviral response induced by double-stranded B-form DNA. *Nature Immunol.* **7**, 40–48 (2006).
- Stetson, D. B. & Medzhitov, R. Recognition of cytosolic DNA activates an IRF3-dependent innate immune response. *Immunity* **24**, 93–103 (2006).
- Guo, W. & Novick, P. The exocyst meets the translocon: A regulatory circuit for secretion and protein synthesis? *Trends Cell Biol.* **14**, 61–63 (2004).
- Lipschutz, J. H., Lingappa, V. R. & Mostov, K. E. The exocyst affects protein synthesis by acting on the translocation machinery of the endoplasmic reticulum. *J. Biol. Chem.* **278**, 20954–20960 (2003).
- Chien, Y. *et al.* RalB GTPase-mediated activation of the I $\kappa$ B family kinase TBK1 couples innate immune signaling to tumor cell survival. *Cell* **127**, 157–170 (2006).
- Balachandran, S. *et al.* Fas-associated death domain-containing protein-mediated antiviral innate immune signaling involves the regulation of Irf7. *J. Immunol.* **178**, 2429–2439 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. Venkataraman, J. Hyun, T. Sato and M. Conkright for technical assistance, M. Gale for RIG-I and IPS-1 lacking MEFs, Y. C. Weh for TBK-1 lacking MEFs, and S. Nagata, T. Maniatis, J. Hiscott and N. Reich for plasmid constructs.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to G.N.B. ([gbarber@med.miami.edu](mailto:gbarber@med.miami.edu)).



## METHODS

**Cells, viruses and reagents.** 293T and HEK293 cells were obtained from the ATCC and were maintained in DMEM medium supplemented with 10% FBS. Human umbilical vein endothelial cells (HUVECs) were obtained from Lonza, and were maintained in endothelial growth medium, EGM-2 (Lonza). *Fadd*<sup>+/+</sup>, *Fadd*<sup>-/-</sup>, *Tbk-1*<sup>+/+</sup> and *Tbk-1*<sup>-/-</sup> MEFs were provided by W.-C. Yeh<sup>16</sup>. VSV-GFP was used in infections and titred as described<sup>16</sup>. VSV-ΔM was constructed as described<sup>17</sup>. Murine IFN-β ELISA Kit was obtained from PBL. EMCV was purchased from ATCC. Synthetic ds B-DNA (poly dA-dT) and poly I:C were purchased from GE Healthcare. Interferon stimulatory DNA (ISD) and reverse sequence of ISD were described previously<sup>20</sup>. For stimulation of cells, B-DNAs or poly I:C were mixed with Lipofectamine 2000 (Invitrogen) at a ratio of 1:1 (vol/wt), and then added to cells at a final concentration of 3 μg ml<sup>-1</sup>.

**Reporter analysis.** 293T cells or MEFs seeded on 24-well plates were transiently transfected with 50 ng of the luciferase reporter plasmid together with a total of 250 ng of various expression plasmids or empty control plasmids. As an internal control, 10 ng of pRL-TK was transfected simultaneously. Then, 36 h later, the luciferase activity in the total cell lysates was measured.

**Northern blot.** Human multiple tissue RNA blots (Clontech) were hybridized with a <sup>32</sup>P-labelled full-length human STING probe.

**Real-time PCR.** Total RNA was isolated from cells using the RNeasy RNA extraction kit (Qiagen) and cDNA synthesis was performed using 1 μg of total RNA (Roche). Fluorescence real-time PCR analysis was performed using a LightCycler 2.0 instrument (Roche Molecular Biochemicals) and TaqMan Gene Expression Assays (Applied Biosystems). Relative amounts of mRNA were normalized to the 18S ribosomal RNA levels in each sample.

**DNA microarray analysis.** Total RNA was extracted from 293T cells transfected with STING expressing vectors. Preparation of cDNA and microarray analysis was performed at the W.M. Keck Foundation Biotechnology Research Laboratory DNA microarray facility at Yale University. The Human Genome U133 Plus 2.0 Array (Affymetrix) was used. Data analysis was performed with GeneSpring software (Silicon Genetics)<sup>16</sup>.

**Mitochondria and ER fraction isolation.** Mitochondria and ER membranes were purified on discontinuous sucrose gradients as previously described<sup>25</sup>. Briefly, HUVEC cells in MTE buffer (0.27 M mannitol, 10 mM Tris-HCl, 0.1 mM EDTA, pH 7.4) were lysed by sonication. Lysed cells were centrifuged at 700g for 10 min to remove nuclei and cellular debris. Mitochondria were obtained by centrifugation at 15,000g for 10 min, and post mitochondrial supernatant was used for purification of ER fractions. Mitochondria pellet was resuspended in MTE buffer, and was layered on discontinuous sucrose gradients consisting of 1.0 M and 1.7 M sucrose and banded by centrifugation at 40,000g for 22 min. Mitochondria fraction was collected and pelleted by centrifugation at 15,000g for 10 min. Purified mitochondria were resuspended in PBS and used for western blot analysis. To isolate ER fractions, postmitochondrial supernatant described above was layered on discontinuous sucrose gradients consisting of 1.3 M, 1.5 M and 2.0 M sucrose, and banded by centrifugation at 100,000g for 70 min. The ER fraction at the interface between the supernatant and the 1.3 M sucrose was collected, and pelleted by centrifugation at 100,000g for 45 min. The ER membranes were resuspended in PBS and were used for western blot analysis.

**RNA interference.** 93T cells were plated on 24-well plates at 5 × 10<sup>4</sup> cells per well and transfected with 10 pmol of siRNA duplex per well using Lipofectamine RNAiMAX (Invitrogen). MEFs were transfected by using an Amaxa nucleofactor apparatus (program A-023) and Amaxa MEF nucleofactor kit 1 according to the manufacturer's recommendations. At 72 h after transfection, cells were used for further experiments.

**Primers.** The following primers were used for cloning: hSTING forward, 5'-CCCAAGCTTGGCCGCCACCATGCCCACTCCAGCCTGC-3'; hSTING reverse,

5'-CCCAAGCTTGGCCGCCACCATGCCCACTCCCACTGCATCC-3'; mSTING forward, CCGCTCGAGAGAGAAATCCGTGCGGAGAG; mSTING reverse, CCGC TCGAGGATGAGGTCA.

The sequences of each siRNA oligonucleotide used in this study are follows: hSTING siRNA, 5'-GCAUCAAGGAUCGGGUUU-3'; mSTING siRNA, 5'-CCAACAGCGUCUACGAGA-3'; SSR2 siRNA, 5'-UCUCAAGGCUUGU AUUU-3', 5'-GUGGAACUAUCUGAUGAU-3', 5'-CAUCUACAAUGUUG GCUC-3', 5'-AAACGAAGAAGAACUGAU-3'; hSEC61B siRNA, 5'-CAGUA UUGGUUAUGAGUC-3'; 5'-GUUCGUAUAUCAGUUAC-3', 5'-GCUCA AAGUUGGCCUGU-3', 5'-CUGUAAGCUUGCUGUUUU-3'; SEC5 siRNA, 5'-CGUCACACCUUCCUAAA-3', 5'-CGGCAUCUCUCCAAAUGA-3', 5'-CAACAGGUGUCAGAAACU-3', 5'-GAAAGGCGGUCUCAGUAC.

**Co-immunoprecipitation, native PAGE and immunoblot analysis.** Cells were seeded on 100-mm dishes at 1 × 10<sup>6</sup> cells per dish. Cells were transfected with a total of 10 μg of empty plasmid or various expression plasmids using Lipofectamine 2000. At 36 h after transfection, cells were lysed in M-PER buffer (Pierce) or 1% digitonin (Calbiochem) buffer (20 mM Tris-HCl, 150 mM NaCl, 1% digitonin) containing protease inhibitors (Roche). Lysates were incubated with HA affinity matrix (Covance) or FLAG affinity beads (Sigma) overnight. The beads were washed three times by TBS containing 0.05% Tween-20 and immunoprecipitates were eluted with non-reducing sample buffer by boiling for 5 min. For endogenous STING immunoprecipitation, HUVECs were lysed in 1% digitonin lysis buffer. Cleared supernatants were incubated with 10 μg of anti-STING antibody, followed by incubation with immobilized protein G (Pierce). The beads were washed four times by 1% digitonin lysis buffer and immunoprecipitates were eluted with SDS sample buffer by boiling for 5 min. Native PAGE and immunoblotting was carried out as previously described<sup>24</sup>.

**Generation of the *Sting* knockout mice.** The linearized targeting vector was electroporated into E14.1 ES cells originated from 129SvEv strain, followed by the selection in G418. Targeted clones were screened by PCR. From 90 clones, 1 positive clone was identified. This ES clone was subjected to the generation of chimaera mice by injection using C57BL/6J blastocysts as the host. The resulting male chimaeras were further mated with C57BL/6J female mice for germline transmission. The heterozygous mice (F<sub>1</sub> mice) were interbred to obtain wild-type, heterozygous and homozygous littermates (F<sub>2</sub>). The genotypes of the mice were determined by PCR. Animals were bred at the University of Miami School of Medicine Transgenic Core Facility. Mice were allowed to freely access to food and water and housed at an ambient temperature of 23 °C and at a 12 h light/dark cycle. Animal care and handling was performed as per IACUC guidelines.

**IFN induced library construction.** To screen for interacting partners of STING, a yeast two hybrid approach was adopted using a novel IFN-induced, telomerase immortalized human fibroblast cDNA library. To construct this library, cells were treated overnight with 1,000 units each of human IFN-α and -β to induce interferon dependent genes. Poly(A)<sup>+</sup> RNA was extracted and cDNA synthesis was carried out using BD Powerscript RT (Clontech). The cDNA was then amplified by PCR, digested with *Sfi*I and ligated to the yeast prey vector pGADT7-RecAB. The ligated mixture was transformed into *Escherichia coli* strain DH10B. The number of independent clones in the unamplified library was estimated to be 3.2 × 10<sup>6</sup>, with an average size of 1.52 kb and inserts ranging in size from 0.8 to 3.0 kb. 15 independent colonies were isolated and DNA extracted to check for size by PCR.

25. Mavinakere, M. S. *et al.* Processing of human cytomegalovirus UL37 mutant glycoproteins in the endoplasmic reticulum lumen prior to mitochondrial importation. *J. Virol.* **80**, 6771–6783 (2006).

# Modelling Myc inhibition as a cancer therapy

Laura Soucek<sup>1</sup>, Jonathan Whitfield<sup>1</sup>, Carla P. Martins<sup>1</sup>, Andrew J. Finch<sup>1</sup>, Daniel J. Murphy<sup>1</sup>, Nicole M. Sodir<sup>1</sup>, Anthony N. Karnezis<sup>1</sup>, Lamorna Brown Swigart<sup>1</sup>, Sergio Nasi<sup>2</sup> & Gerard I. Evan<sup>1</sup>

Myc is a pleiotropic basic helix–loop–helix leucine zipper transcription factor that coordinates expression of the diverse intracellular and extracellular programs that together are necessary for growth and expansion of somatic cells<sup>1</sup>. In principle, this makes inhibition of Myc an attractive pharmacological approach for treating diverse types of cancer. However, enthusiasm has been muted by lack of direct evidence that Myc inhibition would be therapeutically efficacious, concerns that it would induce serious side effects by inhibiting proliferation of normal tissues, and practical difficulties in designing Myc inhibitory drugs. We have modelled genetically both the therapeutic impact and the side effects of systemic Myc inhibition in a preclinical mouse model of Ras-induced lung adenocarcinoma by reversible, systemic expression of a dominant-interfering Myc mutant. We show that Myc inhibition triggers rapid regression of incipient and established lung tumours, defining an unexpected role for endogenous Myc function in the maintenance of Ras-dependent tumours *in vivo*. Systemic Myc inhibition also exerts profound effects on normal regenerating tissues. However, these effects are well tolerated over extended periods and rapidly and completely reversible. Our data demonstrate the feasibility of targeting Myc, a common downstream conduit for many oncogenic signals, as an effective, efficient and tumour-specific cancer therapy.

Myc is deregulated and overexpressed in most cancer cells, where it hijacks the diverse intracellular and extracellular regenerative programs that drive normal cell expansion. Consistent with this, de-activation of Myc in established, Myc-induced transgenic tumours triggers proliferative arrest and re-differentiation of tumour cells, and collapse of the tumour microenvironment and vasculature, usually resulting in rapid tumour regression<sup>2–7</sup>. Although this indicates that Myc might be a good therapeutic target, there are caveats. First, Myc exerts its biological influence through protein–protein and protein–DNA interactions that have proven difficult to disrupt with small molecules. Second, aberrant Myc expression in most human cancers is usually not due to mutation in the *Myc* gene itself but a consequence of its induction by ‘upstream’ oncogenic signals. The therapeutic utility of inhibiting Myc when its aberrant expression is a consequence, not a cause, of oncogenesis is unclear. Finally, Myc is essential for proliferation and stem cell compartment maintenance of regenerative adult tissues such as the gastrointestinal tract, skin and bone marrow. Hence, blocking Myc function systemically might trigger devastating and irreversible side effects. Together, such concerns greatly undermine the credibility of Myc inhibition as an anti-cancer strategy.

Myc-dependent transactivation requires heterodimerization with its bHLHZip partner protein Max<sup>8–10</sup>. Dimerization with Max is also essential for Myc proliferative and oncogenic functions<sup>8,11</sup> and its inhibition has been shown to have potential therapeutic value<sup>12,13</sup>. To model both the therapeutic impact and side effects of Myc inhibition *in vivo*, we constructed a mouse in which endogenous Myc

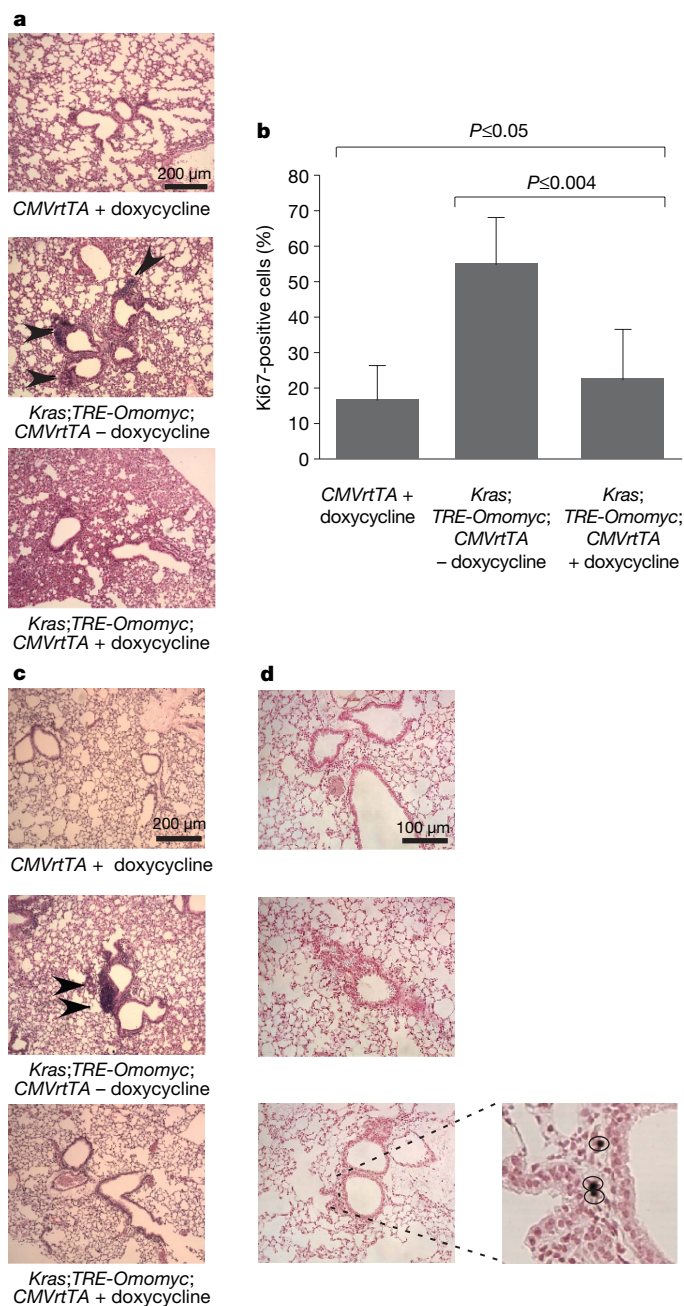
function may be systemically and reversibly inhibited in tissues of adult animals through inducible expression of the dominant interfering Myc bHLHZip dimerization domain mutant Omomyc. Omomyc has four designed amino acid substitutions that facilitate homodimerization with all three oncogenic Myc proteins (c-Myc, N-Myc and L-Myc) but result in little, if any, interaction with Max proteins. Omomyc dimerization denies Max access to Myc<sup>14</sup> and, because Myc–Omomyc heterodimers cannot bind Myc–Max E-box consensus recognition elements, it efficiently blocks Myc-dependent transcriptional activation<sup>15</sup>. Omomyc expression reverses Myc-induced transformation *in vitro*<sup>15</sup> and Myc-driven tumorigenesis *in vivo*<sup>16</sup>.

We conditionally expressed Omomyc using the highly promiscuous cytomegalovirus (CMV) early promoter, which drives relatively high levels of expression in multiple tissue types<sup>17–19</sup>. The *Omomyc* coding sequence was placed downstream of a tetracycline-responsive promoter element (*TRE*) and mice harbouring the *TRE-Omomyc* transgene were then crossed into a background expressing the rtTA transactivator from the CMV early promoter<sup>20</sup>. Omomyc expression was then induced in *TRE-Omomyc;CMVrtTA* double transgenic animals by administration of doxycycline in their drinking water. Real-time polymerase chain reaction with reverse transcription (RT–PCR) confirmed that *Omomyc* expression was detectable in all tested tissues of *TRE-Omomyc;CMVrtTA* double transgenic mice on doxycycline treatment, and was absent from both *TRE-Omomyc* and *CMVrtTA* single transgenic mice, as well as *TRE-Omomyc;CMVrtTA* double transgenic mice not treated with doxycycline (data not shown). Despite some variation between different tissues, steady-state *Omomyc* mRNA levels were demonstrably higher than those of endogenous *c-myc* in most tissues (intestine, kidney, pancreas, heart and lung) and comparable in the rest (for example, spleen and skin) (Supplementary Fig. 1).

To ascertain the therapeutic potential of inhibiting endogenous Myc function in cancer, we chose a well-characterized and validated lung cancer mouse model in which tumorigenesis is initiated by oncogenic activation of endogenous *Kras*<sup>21,22</sup>, a common mutation in human non-small-cell lung cancers. *Kras* is activated in this *LSL-Kras<sup>G12D</sup>* model by inhalation of adenovirus expressing Cre recombinase, which sporadically excises a transcriptional stop element in bronchioalveolar duct junction (BADJ) epithelial cells, triggering expression of oncogenic *Kras<sup>G12D</sup>* driven from the endogenous *Kras* promoter<sup>22</sup>. Tumorigenesis involves development of hyperplasia/adenoma 4–6 weeks after *Kras* activation followed by sporadic progression of some lesions to adenocarcinoma. The resulting mouse lung tumours closely resemble their human counterparts<sup>22,23</sup>. To assess the contribution of endogenous Myc to the initiation, early and late phases of *Kras*-induced lung tumorigenesis, we crossed switchable *TRE-Omomyc;CMVrtTA* mice into the heterozygous *LSL-Kras<sup>G12D</sup>* lung cancer model. To ascertain whether inhibition of Myc prevents initiation of *Kras<sup>G12D</sup>*-induced lung tumours,

<sup>1</sup>Department of Pathology and Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, California 94143-0875, USA. <sup>2</sup>Istituto di Biologia e Patologia Molecolari, C.N.R., University La Sapienza, 00185 Rome, Italy.

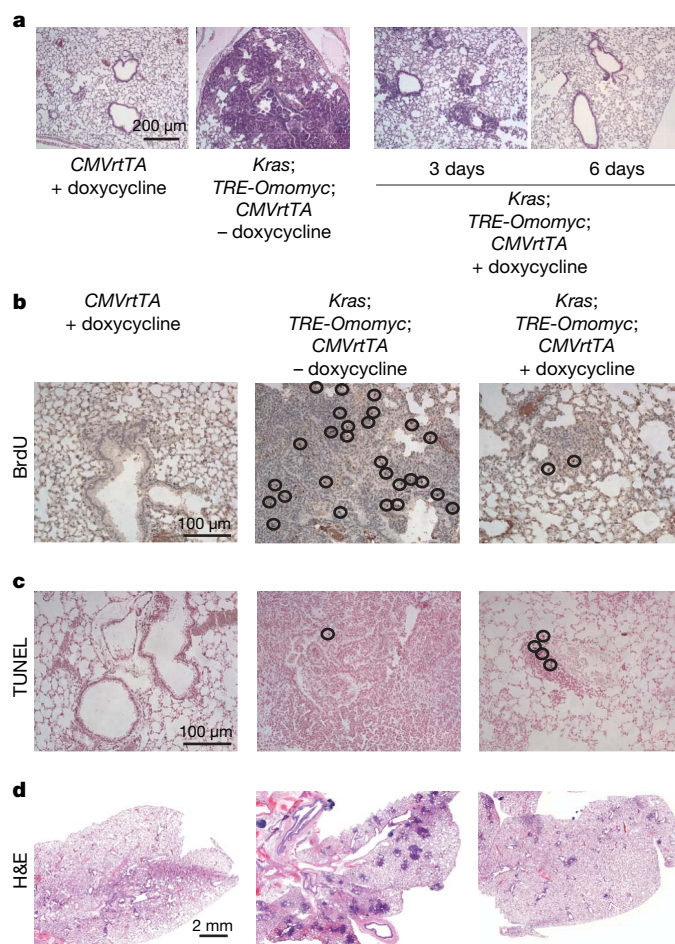




**Figure 1 | Endogenous Myc function is required for formation and maintenance of early-stage *Kras*-induced lung hyperplasias/adenomas.** **a**, Representative haematoxylin-and-eosin-stained sections of lungs from CMVrtTA single transgenic mice (CMVrtTA + doxycycline), untreated *Kras*; *TRE-Omomyc*; *CMVrtTA* triple transgenic mice (*Kras*; *TRE-Omomyc*; *CMVrtTA* – doxycycline) and doxycycline-treated LSL-*Kras*; *TRE-Omomyc*; *CMVrtTA* triple transgenic mice (*Kras*; *TRE-Omomyc*; *CMVrtTA* + doxycycline), 4 weeks after infection with adenoviral Cre. Hyperplastic lesions are indicated by black arrowheads. **b**, Graphical representation of total BADC cells scored as proliferating (Ki67-positive). Error bars represent standard deviation derived from approximately 100 BADCs per mouse. At least three mice were used per series. **c**, Myc inhibition triggers regression of early-stage lung adenomas. Haematoxylin and eosin staining of lungs from mice treated or not with doxycycline for 1 week, starting 6 weeks after Cre-recombinase-expressing adenovirus infection, is shown. A small adenoma is indicated by black arrowheads. **d**, TUNEL staining reveals positive cells in Omomyc and Ras co-expressing samples (*Kras*; *TRE-Omomyc*; *CMVrtTA* + doxycycline), but not in untreated or single transgenic controls (*Kras*; *TRE-Omomyc*; *CMVrtTA* – doxycycline and CMVrtTA + doxycycline). Higher magnification of positive cells is shown in the insert on the right.

680

$5 \times 10^7$  plaque-forming units (p.f.u.) of Cre-recombinase-expressing adenovirus was administered intranasally to LSL-*Kras*; *TRE-Omomyc*; *CMVrtTA* mice and Omomyc induced after 24 h and thereafter maintained for 4 weeks. Mice were then killed and lung morphology, cell proliferation and apoptosis assessed immunohistochemically. For simplicity, only representative controls are shown: doxycycline-treated CMVrtTA mice exemplify all strains lacking an activatable *Kras*<sup>G12D</sup> allele, whereas untreated LSL-*Kras*; *TRE-Omomyc*; *CMVrtTA* mice represent all strains of mice with activated *Kras*<sup>G12D</sup> but no Omomyc. All mice expressing *Kras*<sup>G12D</sup> without Omomyc (Fig. 1a, LSL-*Kras*; *TRE-Omomyc*; *CMVrtTA* – doxycycline) developed multiple hyperplasias and atypical



**Figure 2 | Myc inhibition elicits regression of established lung tumours.** **a**, Haematoxylin-and-eosin-stained lungs from control CMVrtTA single transgenic mice (CMVrtTA + doxycycline), *Kras*<sup>G12D</sup>-expressing mice (*Kras*; *TRE-Omomyc*; *CMVrtTA* – doxycycline) and 3- or 6-day-treated *Kras*<sup>G12D</sup> and Omomyc co-expressing mice (*Kras*; *TRE-Omomyc*; *CMVrtTA* + doxycycline) 18 weeks after infection with adenovirus expressing Cre recombinase. An example of a frank tumour is shown from a mouse expressing *Kras*<sup>G12D</sup> only (*Kras*; *TRE-Omomyc*; *CMVrtTA* – doxycycline). **b**, BrdU staining shows a considerable reduction in BrdU-positive cells in *Kras*<sup>G12D</sup> and Omomyc co-expressing mice (*Kras*; *TRE-Omomyc*; *CMVrtTA* + doxycycline) compared with tissues from mice expressing *Kras*<sup>G12D</sup> only (*Kras*; *TRE-Omomyc*; *CMVrtTA* – doxycycline). **c**, TUNEL staining indicates the presence of apoptotic cells in Omomyc and *Kras*<sup>G12D</sup> co-expressing sections (*Kras*; *TRE-Omomyc*; *CMVrtTA* + doxycycline), but not in untreated or single transgenic control tissues (*Kras*; *TRE-Omomyc*; *CMVrtTA* – doxycycline and CMVrtTA + doxycycline). **d**, Haematoxylin and eosin staining of lung tissue from mice treated for 4 weeks with doxycycline shows clearance of tumour lesions as a consequence of Omomyc expression (compare *Kras*; *TRE-Omomyc*; *CMVrtTA* + doxycycline with *Kras*; *TRE-Omomyc*; *CMVrtTA* – doxycycline).



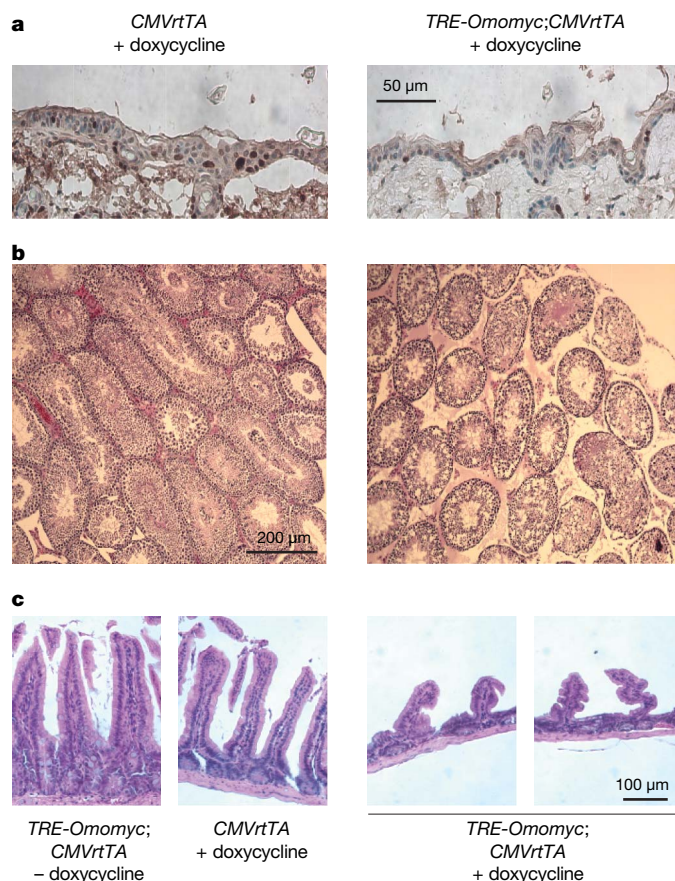
adenomatous hyperplasias (present in  $33 \pm 5.2\%$  of total scored BADJs). By contrast, lesions were absent from all mice expressing *Kras*<sup>G12D</sup> together with Omomyc (*LSL-Kras;TRE-Omomyc;CMVrtTA* + doxycycline). Overall, Omomyc expression significantly suppressed the fraction of cells proliferating (Ki67-positive) at the BADJs (Fig. 1b), consistent with the notion that endogenous Myc is required for proliferation of *Kras*<sup>G12D</sup>-induced early-stage tumour cells. No apoptotic (TdT-mediated dUTP nick end labelling (TUNEL)-positive) cells were present in the BADJs of lungs from mice in any of the test or control groups (data not shown).

To assess the impact of blocking endogenous Myc on maintenance of early-stage Ras-induced lung adenomas, *Kras*<sup>G12D</sup> was activated in *LSL-Kras;TRE-Omomyc;CMVrtTA* mice and nascent tumours allowed to develop for 6 weeks. Representative mice from cohorts of all strains were then killed and examined histologically. All mice harbouring activated *Kras*<sup>G12D</sup> exhibited multiple bronchiolar hyperplasias and adenomatous hyperplasias (present in  $48.7 \pm 6\%$  of total BADJs). Doxycycline was then administered to one-half of the remaining mice for 1 week. Myc inhibition elicited marked shrinkage of lesions and reduction in their multiplicity (identifiable lesions were observed only in  $1.7 \pm 0.8\%$  of BADJs) (Fig. 1c). Consistent with the involution of adenomas, we noted significant apoptosis in the residual lesions of mice expressing Omomyc ( $5.7 \pm 2.5\%$  of total cells). Apoptosis was negligible in control *LSL-Kras;TRE-Omomyc;CMVrtTA* mice in which Omomyc expression had not been induced (0.001%; Fig. 1d). As in the above prevention study, Myc inhibition profoundly reduced the overall numbers of proliferating cells at the BADJs (data not shown).

The induction of apoptosis by Myc inhibition in early-stage *Kras*-induced lung adenomas suggests that, in addition to proliferation, endogenous Myc is required to maintain survival of early-stage *Kras*<sup>G12D</sup>-induced lung tumours. To ascertain whether the same holds true for 'advanced' *Kras*-induced tumours, *Kras*<sup>G12D</sup> was sporadically activated in lungs of *LSL-Kras;TRE-Omomyc;CMVrtTA* mice and tumours allowed to evolve for 18 weeks. At this time, all representative animals exhibited many neoplastic lesions (present in  $89 \pm 6.2\%$  of total BADJs), with some large and highly vascularized adenocarcinomas, as described<sup>22</sup>. Endogenous Myc function was then inhibited in one-half of each cohort and groups of three mice each killed 3, 6 and 28 days later. After only 3 days of sustained Myc inhibition, histological analysis of lungs revealed marked shrinkage of tumours (Fig. 2a). By 6 days we noted a significant reduction in the proliferation of bronchioalveolar epithelial cells ( $1.7 \pm 0.7\%$  5-bromodeoxyuridine (BrdU)-positive cells in BADJs of mice co-expressing Omomyc and *Kras*<sup>G12D</sup> versus  $4.8 \pm 2.2\%$  BrdU-positive cells in mice expressing activated *Kras*<sup>G12D</sup> alone, and  $0.9 \pm 0.3\%$  BrdU-positive cells in *Kras*<sup>G12D</sup>-negative mice; Fig. 2b) accompanied by a significant increase in expression of the senescence marker SA- $\beta$ -galactosidase (data not shown). In addition, we noted significant apoptosis in regressing tumours of Omomyc-expressing mice ( $5.8 \pm 1.3\%$  TUNEL-positive cells versus  $\sim 0.001\%$  in Omomyc negative controls; Fig. 2c). By 28 days of sustained Myc blockade, lungs of animals appeared to be free of obvious tumours with only scattered foci of atypical cells at the BADJs (Fig. 2d). Moreover, during this extended period of Omomyc expression we did not detect any occasional tumours that failed to regress and/or continued to grow, even though each animal developed multiple (hundreds) discrete tumours. Thus, endogenous Myc function is continuously required for proliferation and survival of incipient, early and advanced *Kras*<sup>G12D</sup>-driven lung tumours *in vivo*. We do not yet know if Myc inhibition eradicates all the tumours. However, experiments are underway to monitor for relapse and, should some or all re-grow, for the susceptibility of resurgent tumours to further rounds of Myc inhibition.

Myc has an essential role in the proliferation of all normal cells, raising the possibility that systemic Myc inhibition would trigger severe side effects, especially in continuously proliferating tissues.

Therefore, to explore the impact of systemic Myc inhibition on normal tissues, Omomyc-inducible double transgenic (*TRE-Omomyc;CMVrtTA*) mice were treated continuously with doxycycline for 4 weeks and monitored throughout. Over this extended period, general health and activity of doxycycline-treated *TRE-Omomyc;CMVrtTA* mice appeared indistinguishable from that of non-Omomyc-expressing controls (Omomyc-negative *TRE-Omomyc* and *CMVrtTA* single transgenic mice), with no significant changes in body weight, general activity or blood chemistry (Supplementary Fig. 2 and Supplementary Table 3). After 4 weeks, animals were killed and organs examined. We observed no discernible histological changes in any adult organs with low proliferative indices (for example, pancreas, kidney, liver, heart or lung) (data not shown). However, tissues exhibiting rapid turnover were significantly affected. The basal layer of skin epidermis exhibited marked reduction of proliferating cells ( $12.6 \pm 3.7\%$  of total cells in Omomyc-expressing skin versus  $28.5 \pm 5.5\%$  of basal layer cells in controls) and thinning of the epidermis (Fig. 3a). Furthermore, hair re-growth after shaving was completely inhibited (Supplementary Fig. 3). Nonetheless, Myc inhibition induced no apoptosis, aberrant differentiation or loss of tissue integrity and mice exhibited no signs of skin tearing, ulceration or evident discomfort, even after 8 weeks of sustained Omomyc expression. In testis, Myc inhibition induced



**Figure 3 | Inhibition of endogenous Myc suppresses proliferation in skin, testis and GI tract.** **a**, Anti-Ki67-stained sections of epidermis from 4-week doxycycline-treated *TRE-Omomyc;CMVrtTA* mice and *CMVrtTA* single transgenic controls. **b**, Anti-Ki67-stained sections of testis from 4-week doxycycline-treated *TRE-Omomyc;CMVrtTA* mice show reduced proliferation in seminiferous tubules of doxycycline-treated *TRE-Omomyc;CMVrtTA* mice compared with Omomyc-negative controls. **c**, Haematoxylin-and-eosin-stained small intestine sections show blunted villi in *TRE-Omomyc;CMVrtTA* mice treated for 4 weeks with doxycycline (*TRE-Omomyc;CMVrtTA* + doxycycline) compared with intestine from untreated *TRE-Omomyc;CMVrtTA* and doxycycline-treated *CMVrtTA* single transgenic (*CMVrtTA* + doxycycline) controls.

dramatic atrophy marked by loss of spermatogonia and spermatocytes (Fig. 3b) and some inhibition of cell proliferation ( $32.7 \pm 1.8\%$  of total luminal cells in cycle in Omomyc-expressing mice versus  $42.3\% \pm 3.7\%$  Ki67-positive cells in control mice). We also observed a significant reduction in proliferation in the intestinal crypts (not shown) and striking attrition of villi in the small intestine (Fig. 3c). As with skin, Myc inhibition induced no increase in apoptosis or perturbation of differentiation, and integrity of the epithelium was maintained. Remarkably, all animals retained normal weight, hydration and blood chemistry, confirming maintenance of adequate intestinal absorption and barrier function against bacterial incursion (Supplementary Fig. 2 and Supplementary Table 1).

Despite its widespread activity in most adult tissues, the CMV early promoter functions only weakly in many haematopoietic lineages<sup>20</sup>, precluding its use in modelling the collateral impact of Myc inhibition on bone marrow and lymphoid function. We therefore made use of a complementary model in which doxycycline-dependent expression of Omomyc is driven from the  $\beta$ -actin promoter, which is ubiquitously active in haematopoietic lineages and the stem cell compartments of most tissues, including skin and bone marrow<sup>24–27</sup>. Induction of Omomyc in *TRE-Omomyc;  $\beta$ -actin-rtTA* mice recapitulated the Omomyc phenotypes in skin, gut and testis (Supplementary Fig. 4 and data not shown) and induced profound inhibition of proliferation in bone marrow (50% reduction of BrdU incorporation after 1 week) plus rapid onset of anaemia and leucopenia (Supplementary Table 2). However, by 2 weeks of sustained Myc inhibition, blood counts had returned close to normal levels, aside from mild polycythaemia (Supplementary Table 2). This recovery coincided with onset of extramedullary haematopoiesis in the spleen,

primarily in the erythroid and megakaryocytic lineages (Supplementary Fig. 4). Throughout, all animals showed no overt signs of ill health or distress.

Myc is crucial for the maintenance of stem cell compartments and the balance between self-renewal and differentiation in multiple tissues, including the skin and intestinal crypts<sup>28,29</sup>. This raised the issue of whether even transient Myc inhibition might trigger irreversible changes in sensitive tissues, undercutting the therapeutic practicability of Myc inhibition therapy. Therefore, to assess the reversibility of Omomyc-induced tissue attrition, *TRE-Omomyc; CMVrtTA* mice were treated with doxycycline for 4 weeks. Doxycycline was then withdrawn to restore endogenous Myc function, and the status of affected tissues was followed over time. Within only 1 week of doxycycline withdrawal, cell proliferation had returned to pre-Omomyc levels in the basal layer of skin epithelium (Fig. 4a), hair re-growth re-initiated, and thereafter mice revealed no discernible deficits in any aspect of skin biology (data not shown). Likewise, testis rapidly recovered its full complement of spermatogenic cells including mature sperm (Fig. 4b). In the small intestine, villus length was restored within only 4–5 days (Fig. 4c). Long-term (~1 yr) observation of such mice has revealed no discernible pathology in any tissue. Thus, although systemic suppression of Myc activity has profound effects on proliferating somatic tissues, these effects are well tolerated, without any discernible negative impact on animal well-being and, moreover, they are completely reversible.

Why Myc inhibition should trigger viable arrest in normal tissues but apoptosis in tumours is unclear. Perhaps sustained flux through endogenous Myc is needed to support the increased metabolic and biosynthetic demands made by oncogenic Ras signalling. Endogenous Myc may, like its oncogenic counterpart, also be essential for maintenance of the tumour microenvironment<sup>30</sup>. Whatever the underlying mechanism, our studies suggest that pharmacological inhibition of Myc offers both specificity and efficacy in the treatment of neoplastic disease.

## METHODS SUMMARY

**Transgene construction and generation of mice.** *Omomyc* cDNA was cloned into the pTRE2 plasmid (Clontech). The *TRE-Omomyc* transgene was microinjected into the male pronucleus of day-1-fertilized (*CBA*  $\times$  *C57BL/6*) F<sub>1</sub> embryos, which were transferred into day-1-plugged pseudopregnant foster mice. *CMVrtTA* mice were obtained from the Jackson Laboratory (Tg(*rtTA*hCMV)4Bjd/J). *LSL-Kras<sup>G12D</sup>* mice were gifts from T. Jacks and the  *$\beta$ -actin-rtTA* mice from S. Artandi. Excision of the STOP element was triggered by infection with a Cre-recombinase-expressing adenovirus ( $5 \times 10^7$  p.f.u.)<sup>22</sup>. Doxycycline ( $2 \text{ mg ml}^{-1}$  plus 5% sucrose) was added to drinking water. All mice were treated in accordance with protocols approved by the Institutional Animal Care and Use Committee at UCSF (IACUC approval number AN076148).

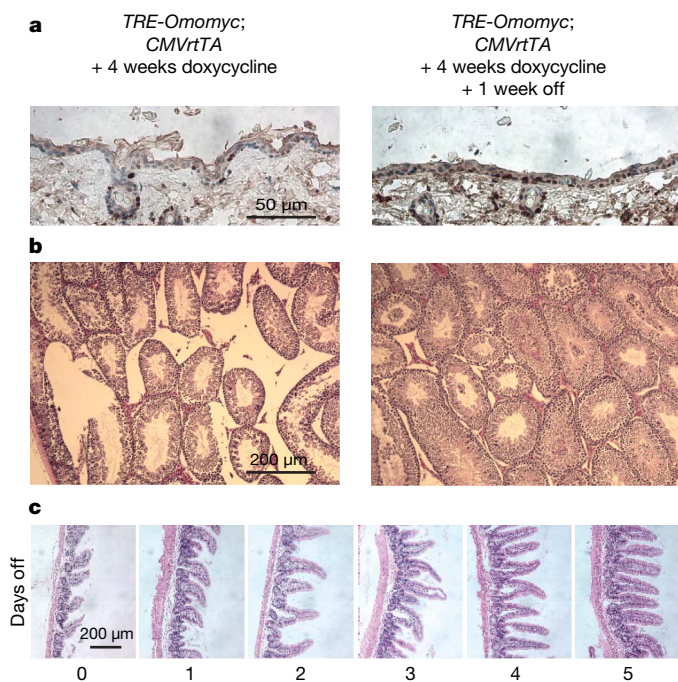
**Histology and immunohistochemistry.** Tissues were stained with haematoxylin and eosin. For antigen retrieval, sections were boiled for 1 min in 0.01 M citrate buffer (pH 6.0). Primary antibody (anti-Ki67 (Neomarkers)) was applied in blocking buffer for 2 h. Incorporated BrdU and apoptotic cells were identified with the BrdU Detection kit II (Roche) and Apoptag kit (Chemicon). HRP-conjugated secondary antibodies (Dako and Molecular Probes) were applied for 30 min and visualized with DAB (Vector Laboratories). Cell proliferation was quantified by scoring Ki67-positive cells as a proportion of total cells at the BADJs in lung tissue (approximately 100 BADJs per mouse), as a percentage of total basal cells in skin (at least 600 total cells per mouse), and as a proportion of total cells in each seminiferous tubule in testis (approximately 80 tubules per section). At least three mice were used for every experimental condition and genotype (Supplementary Table 3).

**Determination of mRNA levels using real-time PCR.** Real-time quantitative RT-PCR (TaqMan) was used to quantify mRNA levels. Total RNA was isolated using Trizol reagent (Gibco). Total RNA ( $1 \mu\text{g}$ ) was reverse transcribed using the iScript cDNA synthesis kit (Bio-Rad). The *Gus* gene was used as an internal amplification control.

Received 1 March; accepted 11 July 2008.

Published online 17 August 2008.

- Oster, S. K., Ho, C. S., Soucie, E. L. & Penn, L. Z. The myc oncogene: Marvelously Complex. *Adv. Cancer Res.* **84**, 81–154 (2002).



**Figure 4 | The degenerative phenotypes induced by systemic Myc inhibition are rapidly and completely reversible on restoration of Myc function.** Myc transactivation function was blocked for 4 weeks by sustained Omomyc expression. Doxycycline was then withdrawn for 1 week. **a**, Ki67 staining indicates rapid recovery of cell proliferation in skin. **b**, Rapid recovery of spermatogenesis in seminiferous tubules. Haematoxylin-and-eosin-stained sections from Omomyc-expressing testis (left panel) and the same tissue 1 week after discontinuing Omomyc expression (right panel) are shown. **c**, Rapid recovery of intestinal villus architecture in *TRE-Omomyc;CMVrtTA* mice after restoration of Myc function. After 4 weeks of sustained Omomyc expression, doxycycline treatment was discontinued and cohorts of mice killed each day for 5 days after doxycycline withdrawal. Representative haematoxylin-and-eosin-stained sections are shown for each time point.



2. Arvanitis, C. & Felsher, D. W. Conditionally MYC: insights from novel transgenic models. *Cancer Lett.* **226**, 95–99 (2005).
3. Felsher, D. W. & Bishop, J. M. Reversible tumorigenesis by MYC in hematopoietic lineages. *Mol. Cell* **4**, 199–207 (1999).
4. Flores, I. *et al.* Defining the temporal requirements for Myc in the progression and maintenance of skin neoplasia. *Oncogene* **23**, 5923–5930 (2004).
5. Jain, M. *et al.* Sustained loss of a neoplastic phenotype by brief inactivation of MYC. *Science* **297**, 102–104 (2002).
6. Pelengaris, S., Khan, M. & Evan, G. I. Suppression of Myc-induced apoptosis in beta cells exposes multiple oncogenic properties of Myc and triggers carcinogenic progression. *Cell* **109**, 321–334 (2002).
7. Pelengaris, S. *et al.* Reversible activation of c-Myc in skin: induction of a complex neoplastic phenotype by a single oncogenic lesion. *Mol. Cell* **3**, 565–577 (1999).
8. Amati, B., Littlewood, T. D., Evan, G. I. & Land, H. The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max. *EMBO J.* **12**, 5083–5087 (1993).
9. Ferre-D'Amare, A. R., Prendergast, G. C., Ziff, E. B. & Burley, S. K. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **363**, 38–45 (1993).
10. Nair, S. K. & Burley, S. K. Structural aspects of interactions within the Myc/Max/Mad network. *Curr. Top. Microbiol. Immunol.* **302**, 123–143 (2006).
11. Amati, B. *et al.* Oncogenic activity of the c-Myc protein requires dimerisation with Max. *Cell* **72**, 233–245 (1993).
12. Chen, J. *et al.* Effects of the MYC oncogene antagonist, MAD, on proliferation, cell cycling and the malignant phenotype of human brain tumour cells. *Nature Med.* **1**, 638–643 (1995).
13. Prochownik, E. V. c-Myc as a therapeutic target in cancer. *Expert Rev. Anticancer Ther.* **4**, 289–302 (2004).
14. Soucek, L. *et al.* Design and properties of a Myc derivative that efficiently homodimerizes. *Oncogene* **17**, 2463–2472 (1998).
15. Soucek, L. *et al.* Omomyc, a potential Myc dominant negative, enhances Myc-induced apoptosis. *Cancer Res.* **62**, 3507–3510 (2002).
16. Soucek, L., Nasi, S. & Evan, G. I. Omomyc expression in skin prevents Myc-induced papillomatosis. *Cell Death Differ.* **11**, 1038–1045 (2004).
17. Baskar, J. F. *et al.* The enhancer domain of the human cytomegalovirus major immediate-early promoter determines cell type-specific expression in transgenic mice. *J. Virol.* **70**, 3207–3214 (1996).
18. Furth, P. A. *et al.* The variability in activity of the universally expressed human cytomegalovirus immediate early gene 1 enhancer/promoter in transgenic mice. *Nucleic Acids Res.* **19**, 6205–6208 (1991).
19. Kothary, R. *et al.* Unusual cell specific expression of a major human cytomegalovirus immediate early gene promoter-lacZ hybrid gene in transgenic mouse embryos. *Mech. Dev.* **35**, 25–31 (1991).
20. Zhan, Y., Brady, J. L., Johnston, A. M. & Lew, A. M. Predominant transgene expression in exocrine pancreas directed by the CMV promoter. *DNA Cell Biol.* **19**, 639–645 (2000).
21. Jackson, E. L. *et al.* The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res.* **65**, 10280–10288 (2005).
22. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **15**, 3243–3248 (2001).
23. Sweet-Cordero, A. *et al.* An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nature Genet.* **37**, 48–55 (2005).
24. Okabe, M. *et al.* 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* **407**, 313–319 (1997).
25. Sarin, K. Y. *et al.* Conditional telomerase induction causes proliferation of hair follicle stem cells. *Nature* **436**, 1048–1052 (2005).
26. Sawamura, D. *et al.* Promoter/enhancer cassettes for keratinocyte gene therapy. *J. Invest. Dermatol.* **112**, 828–830 (1999).
27. Wright, D. E. *et al.* Cyclophosphamide/granulocyte colony-stimulating factor causes selective mobilization of bone marrow hematopoietic stem cells into the blood after M phase of the cell cycle. *Blood* **97**, 2278–2285 (2001).
28. Korinek, V. *et al.* Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nature Genet.* **19**, 379–383 (1998).
29. Murphy, M. J., Wilson, A. & Trumpp, A. More than just proliferation: Myc function in stem cells. *Trends Cell Biol.* **15**, 128–137 (2005).
30. Evan, G. I. Can't kick that oncogene habit. *Cancer Cell* **10**, 345–347 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. Jacks and S. Artandi for their gifts of *LSL-Kras<sup>G12D</sup>* and *β-actin-rtTA* mice, respectively. We thank F. Rostker for technical assistance and Y. Yaron and L. Johnson for advice on the *LSL-Kras<sup>G12D</sup>* model and adenovirus inhalation. We thank our laboratory colleagues for their comments and feedback. This study was supported by grant 2R01 CA98018 from the National Cancer Institute (to G.I.E.). S.N. acknowledges support from AIRC, ASI, CNR, MIUR FIRB and FIRS. C.P.M. is a Leukemia and Lymphoma Society Fellow. J.W. acknowledges support from Human Frontier Science Program. This paper is dedicated to the memory of Judah Folkman.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to G.I.E. ([gevan@cc.ucsf.edu](mailto:gevan@cc.ucsf.edu)).



## LETTERS

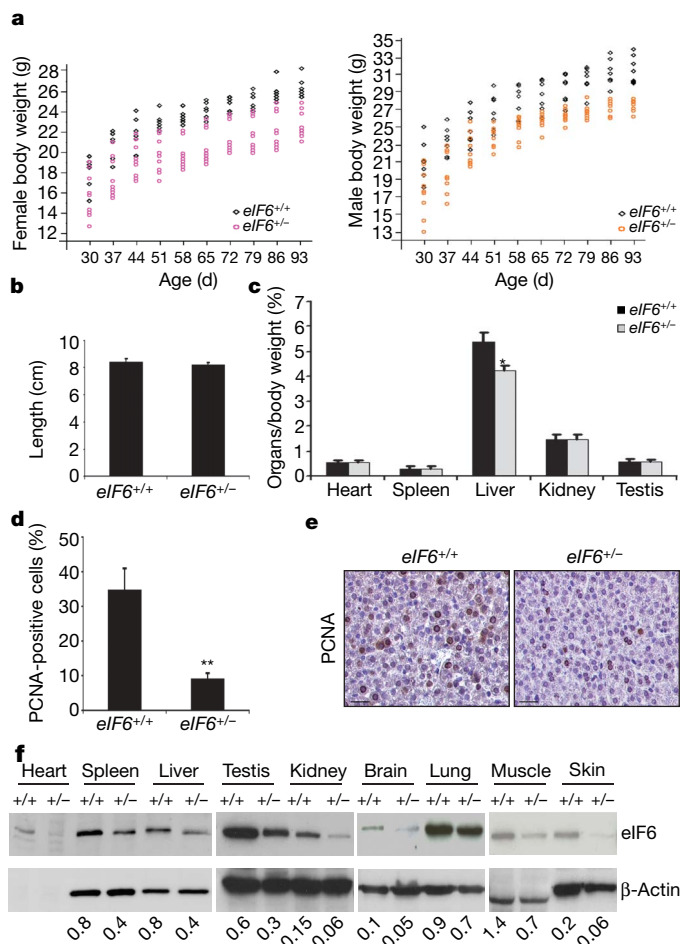
# Eukaryotic initiation factor 6 is rate-limiting in translation, growth and transformation

Valentina Gandin<sup>1</sup>, Annarita Miluzio<sup>1</sup>, Anna Maria Barbieri<sup>1</sup>, Anne Beugnet<sup>1</sup>, Hiroaki Kiyokawa<sup>2\*</sup>, Pier Carlo Marchisio<sup>1,3\*</sup> & Stefano Biffo<sup>1,4</sup>

Cell growth and proliferation require coordinated ribosomal biogenesis and translation. Eukaryotic initiation factors (eIFs) control translation at the rate-limiting step of initiation<sup>1,2</sup>. So far, only two eIFs connect extracellular stimuli to global translation rates<sup>3</sup>: eIF4E acts in the eIF4F complex and regulates binding of capped messenger RNA to 40S subunits, downstream of growth factors<sup>4,5</sup>, and eIF2 controls loading of the ternary complex on the 40S subunit and is inhibited on stress stimuli<sup>6,7</sup>. No eIFs have been found to link extracellular stimuli to the activity of the large 60S ribosomal subunit. eIF6 binds 60S ribosomes precluding ribosome joining *in vitro*<sup>8–10</sup>. However, studies in yeasts showed that eIF6 is required for ribosome biogenesis rather than translation<sup>11–14</sup>. Here we show that mammalian eIF6 is required for efficient initiation of translation, *in vivo*. *eIF6* null embryos are lethal at preimplantation. Heterozygous mice have 50% reduction of eIF6 levels in all tissues, and show reduced mass of hepatic and adipose tissues due to a lower number of cells and to impaired G1/S cell cycle progression. *eIF6*<sup>+/-</sup> cells retain sufficient nucleolar eIF6 and normal ribosome biogenesis. The liver of *eIF6*<sup>+/-</sup> mice displays an increase of 80S in polysomal profiles, indicating a defect in initiation of translation. Consistently, isolated hepatocytes have impaired insulin-stimulated translation. Heterozygous mouse embryonic fibroblasts recapitulate the organism phenotype and have normal ribosome biogenesis, reduced insulin-stimulated translation, and delayed G1/S phase progression. Furthermore, *eIF6*<sup>+/-</sup> cells are resistant to oncogene-induced transformation. Thus, eIF6 is the first eIF associated with the large 60S subunit that regulates translation in response to extracellular signals.

The *eIF6* gene was deleted by homologous recombination using embryonic stem cell technology (Supplementary Fig. 1). The portion of the gene containing the first two exons and the first two introns was substituted by a cassette containing the neomycin resistance gene. The presence of the neomycin resistance cassette did not affect expression of wild-type *eIF6* and of adjacent genes (Supplementary Fig. 2). Germ-line transmission was achieved, and intercrossing of *eIF6* heterozygous mice did not produce any *eIF6*<sup>-/-</sup> offspring. We did not observe *eIF6*<sup>-/-</sup> mice as early as 3.5 days post coitum, indicating that eIF6 is essential for proper development at the stage of preimplantation (Supplementary Table 1). The lethality of *eIF6*<sup>-/-</sup> embryos is consistent with the early expression of the protein in the blastocysts (Supplementary Fig. 1d).

Heterozygous *eIF6*<sup>+/-</sup> mice were viable and indistinguishable from wild-type counterparts up to 30 days after birth. At three months of age, heterozygous mice, independently from gender and genetic background, weighed less than their wild-type littermates (Fig. 1a). The head–anus length of *eIF6*<sup>+/-</sup> and wild-type mice was



**Figure 1 | eIF6 reduction leads to diminished body weight and affects liver growth at the proliferation level.** Age-matched mice were weighed and the growth of organs was followed after birth. **a**, Body weight of females and males recorded from 30 to 93 days after birth. **b**, Length of mice at 93 days, expressed as mean  $\pm$  s.d. ( $n = 8$ ). **c**, The relative weight of indicated organs was calculated as the ratio between organ weight and body weight. Values represent mean  $\pm$  s.d. ( $n = 8$ ). **d**, Percentage of PCNA-positive cells in 9-day-old livers. Values represent mean  $\pm$  s.d. ( $n = 4$ ); **e**, corresponding representative immunohistochemical staining. Scale bar, 100  $\mu$ m. **f**, Representative result of western blot analysis of eIF6 expression in organs from *eIF6*<sup>+/+</sup> and *eIF6*<sup>+/-</sup> adult mice; bottom, densitometric analysis normalized to actin levels. \* $P \leq 0.01$ , \*\* $P \leq 0.05$ .

<sup>1</sup>Molecular Histology and Cell Growth Laboratory, San Raffaele Science Institute, Via Olgettina 58, 20132 Milan, Italy. <sup>2</sup>Northwestern University, Feinberg School of Medicine, 303 East Avenue Chicago, Chicago, Illinois 60611, USA. <sup>3</sup>Università Vita-Salute, Medical School, Via Olgettina 58, 20132 Milan, Italy. <sup>4</sup>DISAV, University of Eastern Piedmont, Via Bellini 8G, 15100 Alessandria, Italy.

\*These authors contributed equally to this work.

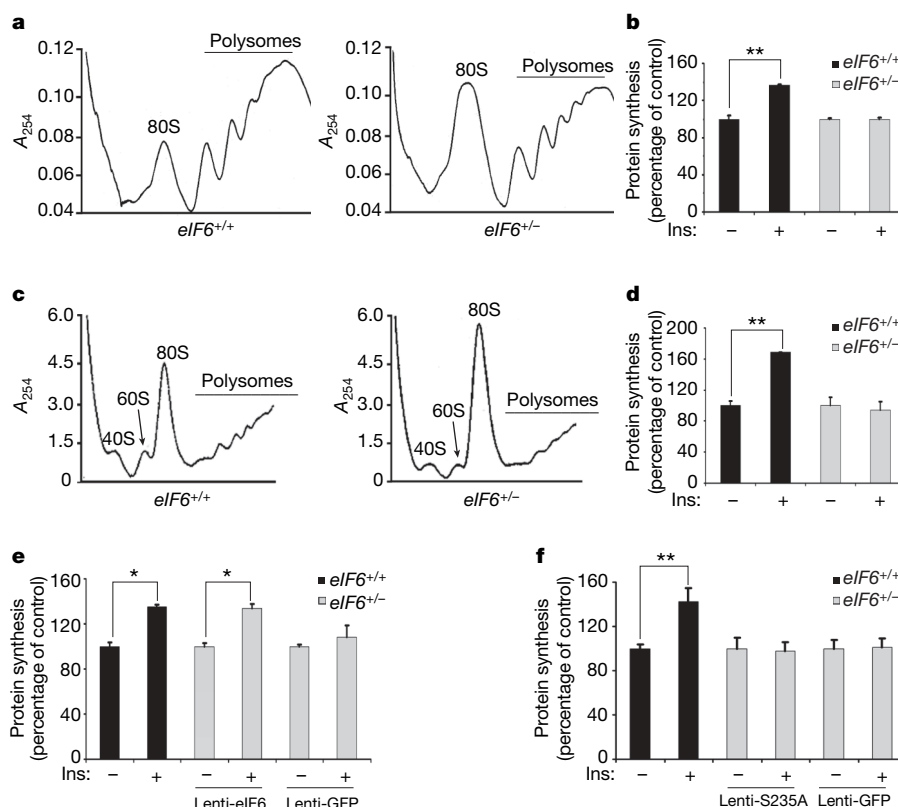
identical (Fig. 1b), suggesting that the reduction of body mass in  $eIF6^{+/-}$  mice could be due to the smaller size of specific organs. Biometric measures of  $eIF6^{+/-}$  and wild-type organs at autopsy revealed that heart and brain were equal to those in wild type, whereas all other organs from  $eIF6^{+/-}$  mice showed a slight trend to mass reduction (not shown). When the weight of the organs was normalized to the body weight of the corresponding animal, for most organs the differences disappeared, whereas the liver (Fig. 1c) and the adipose tissue (Supplementary Fig. 3) exhibited a reduction in their weight/body ratio. A reduction in liver mass could be due to a reduction in cell number, or to a reduction in cell size. Histological staining showed that the liver of  $eIF6^{+/-}$  mice was normal (Supplementary Fig. 4b, d), and further morphometric analysis excluded differences in cell size (Supplementary Fig. 4a, c). In contrast, we observed fewer cycling cells in the liver of  $eIF6^{+/-}$  mice (Fig. 1d, e), suggesting that its reduced mass was due to a reduction in cell number. We performed a similar analysis on adipocytes derived from mesenchymal stem cells: in summary, we found that the reduction of adipose tissue was due to a decreased proliferation of pre-adipocytes, whereas the lipid content per cell was constant (Supplementary Fig. 3). We conclude that the reduction in liver and fat tissue is probably due to a cell autonomous defect marked by reduced proliferative capability.

Quantitative western blotting analysis of eIF6 levels showed that the levels of eIF6 in all heterozygous organs were reduced to half that of the wild type (Fig. 1f; Supplementary Fig. 3c for fat tissue). In spite of reduction of eIF6 levels, we did not detect obvious phenotypes, *in vivo*, in other organs, including the skin (Supplementary Fig. 5), where we reported binding of eIF6 (p27BBP) to  $\beta_4$  integrin<sup>15</sup>. The reduction of eIF6 in all organs from heterozygous mice suggests that under specific circumstances eIF6 haploinsufficiency may become rate-limiting.

We performed polysomal profiles of livers from  $eIF6^{+/-}$  and  $eIF6^{+/+}$  littermates. Approximately 30% of the heterozygous/wild-type couples analysed did not present detectable differences in polysomal profiles. In the remaining 70% heterozygous/wild-type couples, in  $eIF6^{+/-}$  mice polysomal profiles showed an increased 80S peak, accompanied by decreased polysomes (Fig. 2a); this suggests that in mammalian liver  $eIF6^{+/-}$  hepatocytes have a defect in initiation of translation rather than in ribosome biogenesis. We tested this hypothesis using insulin to stimulate basal translation on primary hepatocytes *ex vivo*<sup>16</sup>. We found that insulin administration led to an increase of methionine incorporation in  $eIF6^{+/+}$  hepatocytes, but had little or no effect in  $eIF6^{+/-}$  hepatocytes (Fig. 2b).

We explored further primary mouse embryonic fibroblasts (MEFs) from  $eIF6^{+/-}$  mice. Remarkably, primary MEFs from  $eIF6^{+/-}$  mice showed a phenotype identical to liver hepatocytes with an accumulation of the 80S peak (Fig. 2c) and a reduction in insulin-stimulated translation (Fig. 2d), demonstrating that eIF6 depletion alters translation in different cell types. We previously reported that phorbol 12-myristate 13-acetate (PMA) stimulated methionine incorporation in conditions of eIF6 overexpression<sup>10</sup>. We tested whether PMA-stimulated methionine incorporation was reduced in  $eIF6^{+/-}$  cells, and found that it was (Supplementary Fig. 6). To understand whether basal translation was affected in conditions of eIF6 haploinsufficiency, we pooled data from different experiments. In brief, basal translation is slightly affected in primary hepatocytes, but not in MEFs, consistent with the observation that, *in vivo*, only liver development is affected by eIF6 haploinsufficiency (Supplementary Fig. 7).

We decided to test whether the reduction in methionine incorporation observed in conditions of eIF6 haploinsufficiency could be



**Figure 2 | eIF6 reduction results in 80S complex accumulation and a blunted translational response to insulin.** **a**, Polysomal profiles from livers showing the different amount of 80S ribosomal complex and polysomes between  $eIF6^{+/+}$  and  $eIF6^{+/-}$  *in vivo*. **b**, <sup>35</sup>S-methionine labelling experiments in isolated primary hepatocytes stimulated with 100 nM insulin (Ins). The basal rate of translation is indicated as 100%. Increase of translation is not detected in  $eIF6^{+/-}$  hepatocytes on insulin stimulation.

**c**, Polysomal profile indicates 80S accumulation in heterozygous MEFs; **d**, representative <sup>35</sup>S-methionine labelling of MEFs shows a blunted response to insulin in  $eIF6^{+/-}$ . **e**, **f**, Lentiviral-induced re-expression of wild-type  $eIF6$  in MEFs results in *de novo* insulin stimulation of methionine incorporation (**e**), whereas lentiviral-induced re-expression of  $eIF6^{Ser235Ala}$  does not (**f**). All data represent mean  $\pm$  s.d. of three independent experiments. \* $P \leq 0.01$ , \*\* $P \leq 0.05$ .

due to an indirect inactivation of the mTOR (also known as Frap1) pathway, which regulates translation at the level of initiation. We found that in MEFs, the activation of the PI(3)K–mTOR pathway, measured through the phosphorylation of Akt, rpS6 and 4E-BP1 (also known as Eif4ebp1), was normal (Supplementary Fig. 8). The levels of insulin receptor were also normal (Supplementary Fig. 9).

MEFs reconstituted with eIF6 to wild-type levels of eIF6 (Supplementary Fig. 10) regained insulin responsiveness (Fig. 2e), ruling out indirect effects due to chronic eIF6 depletion. To establish whether translation impairment might be due to the anti-association activity of eIF6, we exploited a Ser235Ala mutant, unable to be released from 60S, *in vitro*, after phosphorylation as we previously described<sup>10</sup>. Heterozygous MEFs reconstituted with eIF6(Ser235Ala) mutant (Supplementary Fig. 10) were not stimulated by insulin and behaved as *eIF6*<sup>+/-</sup> cells (Fig. 2f).

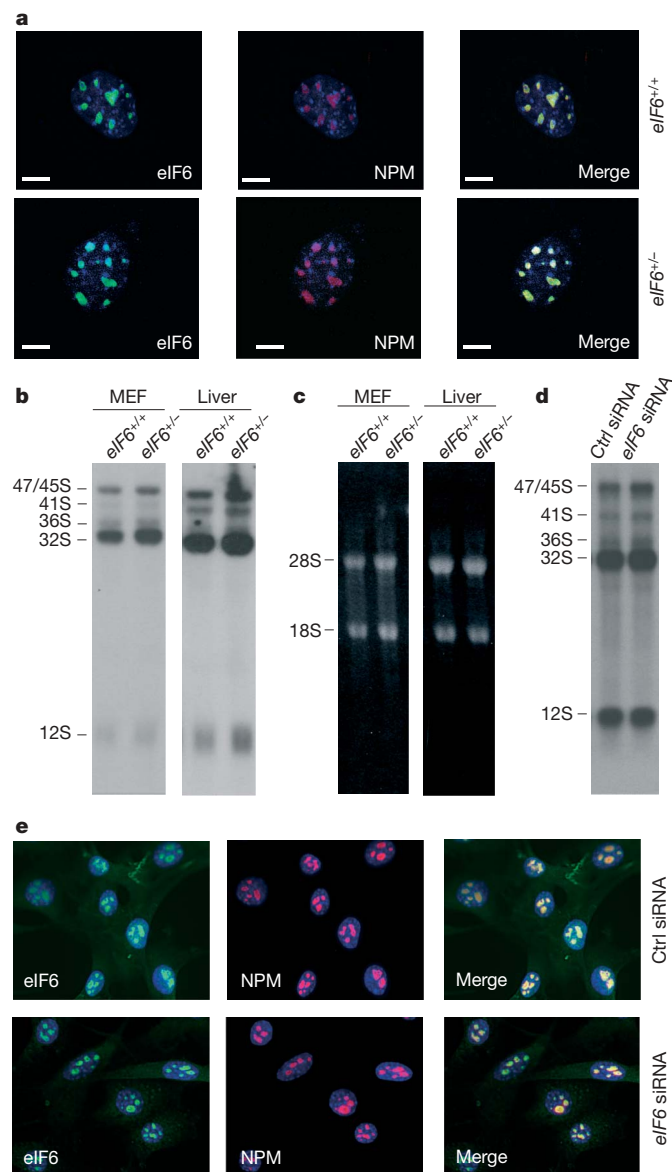
We evaluated in detail whether eIF6 haploinsufficiency can lead to deficits in ribosome biogenesis. In brief, the ratio of free 40S to 60S in *eIF6*<sup>+/-</sup> MEFs was equal to that in *eIF6*<sup>+/+</sup> MEFs, both in the cytoplasm and in the nucleus (Supplementary Fig. 11). The fact that *eIF6*<sup>+/-</sup> MEFs show altered translation and normal ribosome biogenesis can be explained by specific reduction of eIF6 in the cytoplasm, but not in nucleoli. Notably, the reduction of eIF6 that we observed in the cytoplasm of *eIF6*<sup>+/-</sup> cells was more conspicuous than the reduction in the nucleus (Supplementary Fig. 12a). In addition, the (apparent) amount of nucleolar-associated eIF6 was identical in wild-type and heterozygous cells (Fig. 3a).

To investigate whether subtle alterations in ribosome biogenesis could be observed in conditions of haploinsufficiency, we analysed by northern blotting the integrity of the ribosomal RNA processing pathway (Fig. 3b) in MEFs and liver. Our data show that the steady-state level of the rRNA precursors 47/45S, 32S and mature 28S is identical in wild-type and *eIF6*<sup>+/-</sup> cells (Fig. 3b, c). MEFs from early passages (P2/P3) rapidly grow in culture, allowing also for quantitative analysis of rRNA biogenesis<sup>17</sup> by pulse-chase with radio-labelled uridine. We found that *eIF6*<sup>+/-</sup> MEFs process rRNA with normal kinetics (Supplementary Fig. 12c). We also analysed by quantitative polymerase chain reaction with reverse transcription (Q-RT-PCR) the levels of *eIF6* messenger RNA, and two ribosomal proteins rpL5 and rpS19. Our data showed that primary *eIF6*<sup>+/-</sup> MEFs have 50% reduction of *eIF6* mRNA, but no other ribosome-associated proteins, both at the level of mRNA and of protein (Supplementary Fig. 12d, e). Finally, we asked whether a further reduction of eIF6 could lead to a defect in rRNA biogenesis. Eighty per cent depletion of eIF6 by RNA interference in NIH 3T3 cells (Supplementary Fig. 13) showed normal rRNA synthesis (Fig. 3d) and detectable eIF6 in the nucleolus (Fig. 3e). Taken together, these data suggest that the reduction of cytoplasmic eIF6 leads to an increase in the steady-state levels of 80S, accompanied by an impairment of the capability of cells to upregulate protein synthesis on insulin stimulation, but a small amount of eIF6 is sufficient for rRNA biogenesis—in other words, eIF6 is an initiation factor *in vivo*.

We analysed the consequences of eIF6 depletion on cell cycle progression in normal and highly proliferative cells. In early passages, *eIF6*<sup>+/-</sup> MEFs proliferated similarly to wild type, and had equal replicative senescence (Supplementary Fig. 14a, b), rate of apoptosis (Supplementary Fig. 14c, d) and cell size (Supplementary Fig. 14e). In these conditions, cell cycle distribution by fluorescence-activated cell sorting (FACS) analysis of exponentially growing MEFs derived from eIF6 wild type and matched heterozygous did not show gross differences (Fig. 4a). Next, we analysed cell cycle progression in cells synchronized by serum starving and stimulated by serum refeeding. Data obtained by <sup>3</sup>H-thymidine labelling show that *eIF6*<sup>+/-</sup> MEFs enter S-phase 18 h after stimulation (Fig. 4b), compared to the 12 h required by wild-type cells. These data show that *eIF6* haploinsufficiency affects translation and S-phase entry under specific conditions. We looked for additional models in which eIF6 haploinsufficiency could cause a phenotype. eIF6 is overexpressed

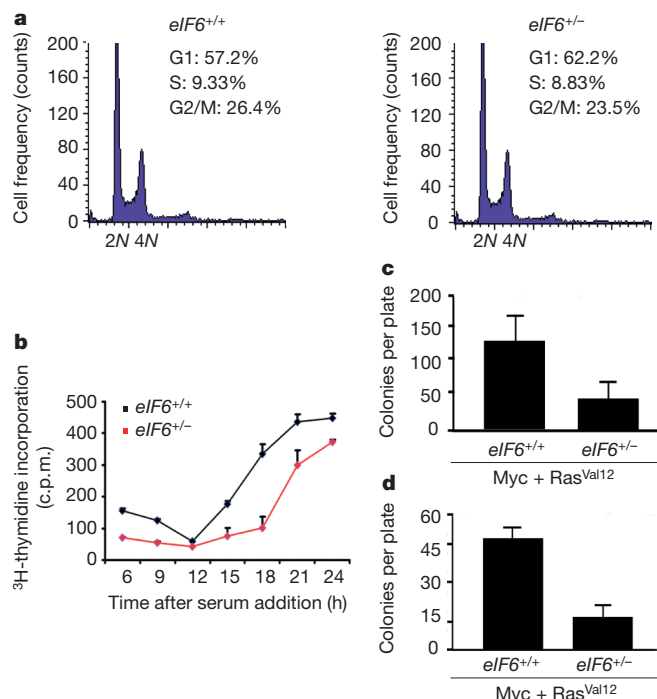
in cancer cells<sup>18,19</sup>. We decided to test whether *eIF6* haploinsufficient MEFs were efficiently transformed by infection with retroviral vectors encoding dominant-negative (DN)p53 and H-ras<sup>V12</sup>, or Myc and H-ras<sup>V12</sup>. *eIF6*<sup>+/-</sup> MEFs formed fewer transformed colonies in soft agar when infected with oncogene-carrying retroviruses compared to wild type (Fig. 4c, d). We had similar results in the focus formation assay (Supplementary Fig. 15a, b). The large difference in the transformation efficiency was not due to differences in retroviral infection and integration (Supplementary Fig. 15c).

Whether eIF6 is a true initiation factor has been unclear for a long time. The effect of eIF6, *in vitro*, ranges from no effects on translation to mild stimulation at low doses accompanied by repression at high doses<sup>8–10</sup>. The difference in the results and in the systems used did not allow a final conclusion. Tif6 (yeast homologue of mammalian eIF6)



**Figure 3 | eIF6 reduction in the cytoplasmic compartment does not affect 60S ribosomal biogenesis.** **a**, eIF6 distribution analysed in the nucleolus by anti-eIF6 immunostaining of primary MEFs. Nucleolar staining with anti-nucleophosmin (NPM) merged with anti-eIF6 (Merge) shows nucleolar eIF6 in *eIF6*<sup>+/-</sup> MEFs compared to wild-type cells. Scale bar, 3  $\mu$ m. **b, c**, Northern blot analysis of steady-state levels of 25S precursors in heterozygotes. MEFs and livers and corresponding ethidium bromide staining (**c**) showing equal steady-state levels of precursor rRNAs. **d, e**, Small interfering RNA (siRNA) of *eIF6* (80% reduction, see Supplementary Fig. 13) in NIH3T3 abolishes neither rRNA processing (**d**) nor nucleolar staining (**e**).





**Figure 4 | eIF6 reduction impairs G1/S progression in synchronized cells and causes reduced transformation.** **a**, FACS analysis of asynchronous MEFs. No significant differences in cell cycle distribution are observed between *eIF6*<sup>+/+</sup> and *eIF6*<sup>+/-</sup> MEFs. **b**, <sup>3</sup>H-thymidine labelling of synchronized cells shows a delay in S-phase entry. **c–d**, Soft agar assay of asynchronous MEFs transformed with DNp53 plus H-ras<sup>V12</sup> (*n* = 6) and with Myc plus H-ras<sup>V12</sup> (*n* = 6). Number of colonies expressed as mean  $\pm$  s.d. indicates that cellular potential to undergo transformation is impaired in *eIF6*<sup>+/-</sup> MEFs.

depletion led to the loss of 60S biogenesis and rRNA processing defects<sup>11–14</sup>. In yeasts, most of Tif6 seems to be localized in the nucleus<sup>20,21</sup> and in pre-ribosomal particles<sup>22</sup>, thus raising questions regarding the physiological relevance of the regulated eIF6 anti-association activity. The results we obtained in *eIF6* knockout mice demonstrate that eIF6 may be rate-limiting for translation, *in vivo*, where it may behave as a stimulatory translation initiation factor downstream insulin/growth factors. From this perspective, eIF6 is the first 60S associated initiation factor able to modulate translation downstream of extracellular signalling *in vivo*. These data do not contradict a role of eIF6 in ribosome biogenesis, but suggest that ribosome biogenesis requires only sequestration of a (minor) nucleolar pool of eIF6.

It remains to be understood whether eIF6 regulates global translation or specific classes of mRNA. It was recently proposed that eIF6 is the mediator regulating microRNA-mediated repression<sup>23</sup> in *Caenorhabditis elegans* and HeLa cells although not in *Drosophila*<sup>24</sup>. In conditions of *eIF6* heterozygosity we did not observe differences in expression of microRNA targets (Supplementary Fig. 16). Further studies are required.

eIF6 is consistently overexpressed in a variety of cancers including colon<sup>18</sup> carcinomas and leukaemias<sup>19</sup>. In the latter context, eIF6 upregulation was interpreted as a by-product of transformation, namely increased growth demanded of ribosomes and translation. Our data support a model in which the levels of eIF6 may directly affect tumorigenesis. In any case, our observations strengthen the concept that translational control is crucial for malignancy, as it was recently proposed for eIF4E, downstream of the mTOR pathway<sup>5</sup>.

## METHODS SUMMARY

All analyses were performed at least three times on different genetic backgrounds. For the generation of knockout mice, we deleted exons 1 and 2.

Hepatocytes and primary MEFs were prepared from littermates and kept in culture for 3 h and 2–3 weeks, respectively. Infection and rescue by eIF6-lentivirus were as detailed in Methods and Supplementary Information.

Analysis of rRNA biogenesis was performed by sucrose gradient centrifugation of dissociated ribosomes, and by pulse-chase rRNA labelling with <sup>3</sup>H-uridine and northern blotting (Supplementary Information). Analysis of translation was performed by sucrose gradient centrifugation and metabolic labelling with <sup>35</sup>S-methionine<sup>25</sup>.

Blastocysts and MEFs were fixed in 2% paraformaldehyde. Blastocysts were then stained by whole-mount immunofluorescence. Cells were attached on coverslips and stained by indirect immunofluorescence.

Primary fibroblasts were infected with retroviruses carrying Myc plus oncogenic H-ras<sup>V12</sup> or DNp53 plus oncogenic H-ras<sup>V12</sup>. Foci were counted 2–3 weeks after the primary infection<sup>26</sup>.

SDS-PAGE and western blotting analyses were performed on total extracts in RIPA buffer (10 mM Tris-HCl, pH 7.2, 1% sodium deoxycholate, 1% Triton X-100, 0.1% SDS, 150 mM NaCl, 1 mM EDTA, pH 8.0) as detailed elsewhere<sup>17</sup>. Nuclear–cytoplasmic fractionation was performed by ultracentrifugation on a sucrose cushion as described in Supplementary Methods. The purity of the fractions was checked by antibodies against Gapdh (cytoplasm) or histone H2B (also known as Hist2h2be, nucleus). We loaded equivalent amounts of cells.

Cell number and viability were analysed as previously described<sup>27</sup>. S-phase entry was analysed by the <sup>3</sup>H-thymidine assay on primary embryo fibroblasts at early passages.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 26 October 2007; accepted 15 July 2008.

Published online 10 September 2008.

- Kapp, L. D. & Lorsch, J. R. The molecular mechanics of eukaryotic translation. *Annu. Rev. Biochem.* **73**, 657–704 (2004).
- Gebauer, F. & Hentze, M. W. Molecular mechanisms of translational control. *Nature Rev. Mol. Cell Biol.* **5**, 827–835 (2004).
- Proud, C. G. Signalling to translation: how signal transduction pathways control the protein synthetic machinery. *Biochem. J.* **403**, 217–234 (2007).
- Sonenberg, N. & Pause, A. Signal transduction. Protein synthesis and oncogenesis meet again. *Science* **314**, 428–429 (2006).
- Mamane, Y., Petroulakis, E., LeBacquer, O. & Sonenberg, N. mTOR, translation initiation and cancer. *Oncogene* **25**, 6416–6422 (2006).
- Holcik, M. & Sonenberg, N. Translational control in stress and apoptosis. *Nature Rev. Mol. Cell Biol.* **6**, 318–327 (2005).
- Wek, R. C., Jiang, H. Y. & Anthony, T. G. Coping with stress: eIF2 kinases and translational control. *Biochem. Soc. Trans.* **34**, 7–11 (2006).
- Russell, D. W. & Spremulli, L. L. Purification and characterization of a ribosome dissociation factor (eukaryotic initiation factor 6) from wheat germ. *J. Biol. Chem.* **254**, 8796–8800 (1979).
- Valenzuela, D. M., Chaudhuri, A. & Maitra, U. Eukaryotic ribosomal subunit anti-association activity of calf liver is contained in a single polypeptide chain protein of Mr = 25,500 (eukaryotic initiation factor 6). *J. Biol. Chem.* **257**, 7712–7719 (1982).
- Ceci, M. *et al.* Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly. *Nature* **426**, 579–584 (2003).
- Sanvito, F. *et al.* The  $\beta 4$  integrin interactor p27(BBP/eIF6) is an essential nuclear matrix protein involved in 60S ribosomal subunit assembly. *J. Cell Biol.* **144**, 823–837 (1999).
- Si, K. & Maitra, U. The *Saccharomyces cerevisiae* homologue of mammalian translation initiation factor 6 does not function as a translation initiation factor. *Mol. Cell Biol.* **19**, 1416–1426 (1999).
- Wood, L. C., Ashby, M. N., Grunfeld, C. & Feingold, K. R. Cloning of murine translation initiation factor 6 and functional analysis of the homologous sequence YPR016c in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **274**, 11653–11659 (1999).
- Basu, U., Si, K., Warner, J. R. & Maitra, U. The *Saccharomyces cerevisiae* TIF6 gene encoding translation initiation factor 6 is required for 60S ribosomal subunit biogenesis. *Mol. Cell Biol.* **21**, 1453–1462 (2001).
- Biffo, S. *et al.* Isolation of a novel  $\beta 4$  integrin-binding protein (p27(BBP)) highly expressed in epithelial cells. *J. Biol. Chem.* **272**, 30314–30321 (1997).
- Clark, R. L. & Hansen, R. J. Insulin stimulates synthesis of soluble proteins in isolated rat hepatocytes. *Biochem. J.* **190**, 615–619 (1980).
- Strezoska, Z., Pestov, D. G. & Lau, L. F. Bop1 is a mouse WD40 repeat nucleolar protein involved in 28S and 5.8S rRNA processing and 60S ribosome biogenesis. *Mol. Cell Biol.* **20**, 5516–5528 (2000).
- Sanvito, F. *et al.* Expression of a highly conserved protein, p27BBP, during the progression of human colorectal cancer. *Cancer Res.* **60**, 510–516 (2000).
- Harris, M. N. *et al.* Comparative proteomic analysis of all-trans-retinoic acid treatment reveals systematic posttranscriptional control mechanisms in acute promyelocytic leukemia. *Blood* **104**, 1314–1323 (2004).

20. Senger, B. *et al.* The nucleolar Tif6p and Efl1p are required for a late cytoplasmic step of ribosome synthesis. *Mol. Cell* **8**, 1363–1373 (2001).
21. Menne, T. F. *et al.* The Shwachman–Bodian–Diamond syndrome protein mediates translational activation of ribosomes in yeast. *Nature Genet.* **39**, 486–495 (2007).
22. Volta, V. *et al.* Sen34p depletion blocks tRNA splicing *in vivo* and delays rRNA processing. *Biochem. Biophys. Res. Commun.* **337**, 89–94 (2005).
23. Chendrimada, T. P. *et al.* MicroRNA silencing through RISC recruitment of eIF6. *Nature* **447**, 823–828 (2007).
24. Eulalio, A., Huntzinger, E. & Izaurralde, E. GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nature Struct. Mol. Biol.* **15**, 346–353 (2008).
25. Gorrini, C. *et al.* Fibronectin controls cap-dependent translation through  $\beta 1$  integrin and eukaryotic initiation factors 4 and 2 coordinated pathways. *Proc. Natl Acad. Sci. USA* **102**, 9200–9205 (2005).
26. Zou, X. *et al.* Cdk4 disruption renders primary mouse cells resistant to oncogenic transformation, leading to Arf/p53-independent senescence. *Genes Dev.* **16**, 2923–2934 (2002).
27. Yang, Y. L. *et al.* BubR1 deficiency results in enhanced activation of MEK and ERKs upon microtubule stresses. *Cell Prolif.* **40**, 397–410 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by grants AIRC (S.B., P.C.M.), TELETHON GGB05043, CARIPO 0578 (S.B.) and NIH-RO1 (H.K.). A.B. is supported by grant AICR 05-360. The manuscript has been improved thanks to suggestions from N. Offenhaeuser and A. Boletta. We are indebted to P. G. Pelicci for anti-NPM antibodies, H. Hirai for preliminary soft agar assays, S. Modina for blastocyst preparation, M. Vidali for hepatocytes preparation, M. Malosio for insulin receptor antibodies, S. Gregori for FACS analysis, D. Bartel for reporter constructs, G. Manfioletti for HMGA2 antibodies and F. Loreni for rpS19 antibodies. We acknowledge L. Magri for preliminary experiments, and V. Volta and S. Grosso for suggestions.

**Author Contributions** V.G., A.M., A.M.B., H.K. and S.B. planned the experiments; V.G., A.M., A.M.B., A.B. and S.B. performed the experiments; all authors analysed the data; and V.G. and S.B. wrote the paper. All authors discussed the results and contributed to the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.B. ([stefano.biffo@hsr.it](mailto:stefano.biffo@hsr.it)).

## METHODS

**Antibodies and reagents.** The following antibodies were used: rabbit polyclonal against eIF6 (ref. 15), caspase 3 (Cell Signaling), PCNA (Abcam) and H2B (Santa Cruz), and mouse monoclonal against  $\beta$ -actin (Sigma), nucleophosmin<sup>28</sup> and Gapdh (Chemicon International). Nuclear staining was performed with Hoechst 3342 (Molecular Probes). Human insulin was from Sigma.

**Generation of eIF6 null mice.** *eIF6* gene retrieval was obtained by screening a mouse genomic DNA library (SJ129 strain). The gene was cloned into a pPNT plasmid. The 5' long arm spanned to the exon 1 and the 3' short arm included exon 3 and exon 4. The construct was electroporated in SJ129 embryonic stem cells by the Core Facility for Conditional Mutagenesis (CFCM). eIF6 mutant cells obtained by homologous recombination were identified by Southern blot and PCR analysis (Supplementary Methods). Recombined embryonic stem cells were used to produce chimaeric animals by both injection and aggregation techniques. Chimaeras were then mated to SJ129 mice for germline transmission. Backcross was performed to the CL57Bl6 background. Genotyping analysis was performed by PCR on tail genomic DNA (Supplementary Methods). Controls for the proper expression of upstream and downstream genes to *eIF6* (*Itgb4bp*) were performed by RT-PCR as detailed in Supplementary Methods.

All experiments involving animals were performed in accordance with Italian national regulations and experimental protocols reviewed by local Institutional Animal Care and Use Committees (IACUC n. 574).

**Blastocyst preparation and immunofluorescence.** Blastocysts from hyperovulated, mated C57Bl6 mice were obtained by flushing the uterine tubes with M2 medium (Sigma). Whole-mount immunofluorescence was performed as detailed previously<sup>29</sup>. Images were obtained with fluorescent microscopy (Zeiss Axiophot) or laser confocal microscopy (Ultra View, Perkin Elmer). Image slides were taken with a digital camera (Sensicam) using Image-Pro plus software.

Immunofluorescence on MEFs was performed as described previously<sup>11</sup>, after fixation for 10 min at 25 °C in 2% buffered paraformaldehyde, 3% sucrose in PBS, followed by permeabilization in 20 mM Hepes, pH 7.4, 0.5% Triton-X100.

**Polysomal profiles.** Polysomal profiles were performed as described previously<sup>10</sup>. In brief, after measuring  $A_{254}$  of the extracts prepared in 30 mM Tris-HCl, pH 7.5, 100 mM NaCl, 30 mM MgCl<sub>2</sub>, 0.1% NP-40 and 10  $\mu$ g ml<sup>-1</sup> cycloheximide, the equivalent of 5 units of absorbance were layered on a 15–50% sucrose gradient and centrifuged at 4 °C in a SW41Ti Beckman rotor for 3 h 30 min at 39,000 r.p.m., corresponding to 190,000g. Absorbance at 254 nm was recorded by BioLogic LP software (BioRad). The preparation of individual ribosomal subunits and polysomes was obtained as detailed in Supplementary Methods. In brief, ribosomal subunits were dissociated by 20 mM EDTA after isolating them at 39,000 r.p.m. (190,000g) for 16 h on a 30% sucrose cushion then separated on a 10–30% sucrose gradient at 36,000 r.p.m. (160,000g) and analysed as above.

Analysis of rRNA metabolism in primary embryo fibroblasts was performed as detailed in Supplementary Methods.

**Primary cell cultures.** Hepatocytes and primary MEFs were prepared from littermates and kept in culture for 3 h and 2–3 weeks, respectively. All the analyses were performed at least three times on different genetic backgrounds. Infection and rescue by eIF6-lentivirus was as detailed in Supplementary Information.

Primary embryo fibroblasts (MEFs) were prepared from E13.5 embryos and cultured as described<sup>30</sup>. In brief, embryos were dissociated by 0.03% trypsin in 0.2% EDTA at 37 °C for 10 min and then treated with DNaseI (New England Biolabs). After filtering fibroblasts with a 70  $\mu$ m cell strainer, they were cultured in DMEM (Gibco) supplemented with 10% fetal bovine serum (FBS), 100  $\mu$ g ml<sup>-1</sup> streptomycin, 100 U ml<sup>-1</sup> penicillin and 2 mM glutamine. MEFs were cultured at 9% CO<sub>2</sub> at 37 °C and experiments were performed at first passages (P1–P4). Hepatocytes were isolated as detailed in Supplementary Methods.

**Protein synthesis measurement by <sup>35</sup>S-methionine labelling.** Primary embryo fibroblasts (MEFs) and primary hepatocytes were used for analysis of translational rate. Cells were seeded at sub-confluency in 6-well plates, and insulin stimulation was performed at 100 nM for 30 min. Metabolic labelling was performed as described previously<sup>25</sup>. All experiments were done in triplicate. Statistical *P*-values calculated by two-tailed *t*-test were also indicated; two asterisks for *P*-values less than 0.05 and one asterisk for *P*-values less than 0.01.

**Immunohistochemistry and histological staining.** Immunohistochemical and histological analysis were performed on paraffin-embedded sections prepared as detailed in Supplementary Methods.

**Cell cycle analysis.** S-phase entry was analysed by <sup>3</sup>H-thymidine assay on primary embryo fibroblasts at early passages. MEFs were seeded at 60% confluency in 6-well plate and starved for 48 h in DMEM plus 0.1% FBS. MEFs were then

stimulated with 10% serum and labelled with 1  $\mu$ Ci ml<sup>-1</sup> <sup>3</sup>H-thymidine for 24 h. 6, 9, 12, 15, 18, 21 and 24 h after serum addition, <sup>3</sup>H-thymidine was added to the medium and cells were lysed after 1 h as described in Supplementary Methods. Half of the radioactive sample obtained (0.5 ml) was solubilized in scintillation fluid and counted with a  $\beta$ -counter. Data derived from the mean of three independent experiments and counts per minute were normalized to the total amount of DNA.

**Retroviral infection and soft agar assay.** The Phoenix ecotropic virus packaging cells were obtained from the American Tissue Culture Collection (ATCC) and maintained in DMEM supplemented with 10% FBS, 1% glutamine and antibiotics. The pBABE-hygro vector for expression of H-ras<sup>V12</sup>, dominant-negative p53 mutant (DNp53) and Myc was described previously<sup>30</sup>. The soft agar formation assay and the focus formation assay were performed as described previously<sup>30</sup>.

**Southern blot and PCR analysis.** Southern blot hybridization was used to identify the targeted allele in the embryonic stem cells using standard protocols. Specifically, targeted clones were identified by using a 5' external probe, which detects the change of a 16.3 kb wild-type EcoRV fragment to a novel 13 kb EcoRV fragment, and then were confirmed with a 3' external probe, which detects the change of a 13.3 kb wild-type HindIII fragment into a novel 9.2 kb HindIII fragment.

Genotyping of the offspring mice was performed using a three-primers PCR strategy to simultaneously amplify target sequences from both the wild-type and the mutant alleles. PCR was performed by AmpliTaq Gold (Roche) according to the manufacturer's protocol. A shared primer number 3 (5'-GTG AGTCTTGGCTTCATGTG-3') was designed downstream of the deleted region. This primer can pair with wild-type allele-specific primer number 2 (5'-CTATGTGGCCTTGGTCCAC-3') to amplify a 320 bp PCR product or with the targeted allele-specific primer number 1 (5'-GCAGCGATCGCCTTC TATC-3') in the *neo* cassette to amplify a 650 bp fragment from the mutant allele. PCR products were resolved on 2% agarose gel.

**mRNA extraction and real-time RT-PCR.** Total RNA was extracted with TRIzol reagent (Invitrogen) from MEFs and from tested organs of *eIF6*<sup>+/+</sup> and *eIF6*<sup>+/-</sup> mice. After treatment of total RNA with RQ1 RNase-free DNase (Promega), reverse transcription was performed with SuperScript™ First-Strand (Invitrogen) according to the manufacturer's instructions. 100 ng of reverse-transcribed complementary DNA was amplified with the appropriate primers and detected with probes specific for *Itgb4bp*, *Mmp24*, *Gdf5*, *Rpl5* and *Rps19* cDNA (TaqMan gene expression assay Mm00550245\_m1, Mm00487721\_m1, Mm00433564\_m1, Mm00847026\_g1 and Mm00452264\_m1, respectively, from Applied Biosystems) in an ABI PRISM 7900HT Sequence Detection System. *Gapdh* was used as internal control gene expression (TaqMan assay reagent 4352932E). The data were expressed as the percentage of relative quantity of target genes. Results are represented as means of three independent experiments.

**Immunohistochemistry and histological staining.** Tissue samples were excised from *eIF6*<sup>+/+</sup> and *eIF6*<sup>+/-</sup> mice and immediately fixed in 10% formaldehyde in 0.1 M phosphate buffer (PBS) at pH 7.4. Serial 3- $\mu$ m sections of paraffin-embedded liver, skin and fat were stained with haematoxylin and eosin (H&E) for morphological analysis. Liver sections were subjected to immunohistochemical staining for proliferating cell nuclear antigen (PCNA) using the Vectastain Elite ABC kit (Vector). In brief, slides were passed in graded alcohol solutions, rinsed in distilled water and immersed in citrate buffer (0.1 M citric acid plus 0.1 M sodium citrate). Samples were boiled in a microwave oven three times for 5 min, washed in PBS and soaked in 3% hydrogen peroxide for 30 min to block endogenous peroxidase activity. After blocking in normal serum, slides were incubated with anti-PCNA (1:2,000) for 16 h at 4 °C. Sections were incubated with a biotinylated secondary antibody for 1 h at room temperature and then with a preformed avidin-biotinylated peroxidase complex. Signals were detected using diaminobenzidine tetrahydrochloride (DAB, BioGENEX) and sections were counterstained with haematoxylin. Stained samples were observed in a Zeiss Axiophot microscope and pictures were acquired with ScionImage Software.

**siRNA transfection of NIH 3T3 cells.** Transient transfection was performed on NIH 3T3 cells using Lipofectamin 2000 (Invitrogen), according to manufacturer's instructions.

For siRNA experiments, siGENOME duplex targeting ITGB4BP was used (NM\_002212; sense sequence, 5'-GAGCUUCGUUCGAGAACAUAU-3'; antisense sequence, 5'-PUUGUUCUGAAGCAGCUCUU-3'; Dharmacon). SiCONTROL NON-TARGETING siRNA POOL (Dharmacon) was used as negative control.

Cells were transfected at 30–40% confluency using 100 pmol siRNA in Opti-MEM medium (Invitrogen); after 3 h at 37 °C the medium was changed, and gene silencing was examined by western blot analysis 72 h after transfection.

**Flow cytometry and measurement of cell size.** Unfixed *eIF6*<sup>+/+</sup> and *eIF6*<sup>+/-</sup> MEFs were pelleted and resuspended in PBS containing 5 ng  $\mu$ l<sup>-1</sup> propidium



iodide (Sigma). Cell size (FSC) was analysed using FACS Canto II (Becton Dickinson).

**Cell fractionation.** Cells were detached by trypsinization and collected by centrifugation. The pellet was resuspended in ice-cold hypotonic wash buffer (10 mM Tris HCl, pH 7.4, 10 mM KCl, 2 mM MgCl<sub>2</sub>) and then the cytoplasmic fraction was extracted in hypotonic lysis buffer (10 mM Tris HCl, pH 7.4, 10 mM KCl, 2 mM MgCl<sub>2</sub>, 0.05% Triton X-100, 1 mM EGTA, 1 mM DTT) for 20 min on ice. Nuclei were collected by centrifugation at 700g for 15 min and resuspended in hypotonic wash buffer before loading on 1.6 M sucrose cushion for nuclear fraction purification. The nuclear pellet was obtained by ultracentrifugation (TL100, Beckman) at 50,000 r.p.m. for 1 h and then lysed in RIPA buffer.

**Lentiviral infection.** 293T cells (ATCC) were infected with packaging plasmid VSV-G, PMDLg/pRRE, pREV and transfer vector pCCL-PPT-hPGK-pre previously described<sup>31</sup>. Full-length eIF6 and GFP were cloned in transfer vector. MEFs were infected with eIF6-pCCL and GFP-pCCL at early passage (P<sub>1</sub>) and used for <sup>35</sup>S-methionine labelling as described. Multiplicity of infection (m.o.i.) was empirically determined to achieve twofold eIF6 expression by triplicate infections.

*In vitro* mutagenesis was performed on the wild-type eIF6 lentiviral transfer plasmid by the QuickChange Kit (Stratagene) using the same oligonucleotides as in ref. 32.

**Ear mesenchymal stem cell culture and adipogenesis.** Mesenchymal stem cells from outer ears of eIF6<sup>+/+</sup> and eIF6<sup>+/-</sup> mice were collected and differentiated into the adipogenic lineage as previously described<sup>33</sup>.

Differentiated cells were fixed in 10% buffered formaline and stained with Oil Red O for 10 min. The dye retained by the cells was eluted with isopropanol and quantified by measuring absorbance at 500 nm.

**Polysomal profile from liver.** Tissue was gently homogenized in 50 mM Tris HCl, pH 7.8, 240 mM KCl, 10 mM MgSO<sub>4</sub>, 5 mM DTT, 250 mM sucrose, 2% Triton X-100, 90 µg ml<sup>-1</sup> cycloheximide and 30 U ml<sup>-1</sup> RNasin using a glass douncer. After clarification, 100 µg ml<sup>-1</sup> heparin was added to the supernatant. 5 OD<sub>254</sub> of rRNA was loaded on a 15–55% sucrose gradient dissolved in 25 mM Tris HCl, pH 7.4, 25 mM NaCl, 5 mM MgCl<sub>2</sub> and 1 mM DTT and run at 39,000 r.p.m. (190,000g) for 3 h 30 min with SW41Ti swing rotor (Beckman Coulter). The polysomal profile was detected as previously described.

**RNA blot analysis.** Total RNA was isolated from MEF cells or from mouse liver using TRI REAGENT (Sigma) following the manufacturer's protocol. RNA (5 µg) was separated on 1% agarose-formaldehyde gel and analysed by northern blot hybridization using standard techniques.

The oligonucleotide ITS2-2 complementary to nucleotides 239–271 of the ITS2 region was used as probe for hybridization (5'-ACTGGTGAGGCA GCGGTCCGGGAGGCGCCGACG-3'), as described previously<sup>34</sup>. The probe was 5'-labelled using [ $\gamma$ -<sup>32</sup>P] ATP and T4 polynucleotide kinase.

**Preparation of 40S and 60S ribosomal subunits.** Dissociation of 40S and 60S ribosomal subunits extracted from cytoplasmic and nuclear compartments were obtained as described previously<sup>20</sup>. 40S and 60S subunits were dissociated with 20 mM EDTA. The equivalent of 20 units of A<sub>254</sub> of total rRNA were loaded on a 30% sucrose cushion dissolved in 10 mM Tris HCl, pH 7.5, 10 mM KCl, 1 mM MgCl<sub>2</sub> and 1 mM DTT. Ribosomal RNA was collected by ultracentrifugation at 36,000 r.p.m. for 16 h, washed twice in cold 1 × PBS and resuspended in 30 mM Tris HCl, pH 7.5, 10 mM KCl, 1 mM MgCl<sub>2</sub>, 20 mM EDTA and 1 mM DTT. 10 OD of dissociated 40S and 60S subunits were separated through a 10–30%

sucrose gradient, ultracentrifuged at 36,000 r.p.m. for 3 h 30 min and analysed by Biological LP System (BioRad) as described previously.

**Hepatocytes isolation.** One-month-old mice were anaesthetized by intraperitoneal injection of 200 µl chloralium hydrate (8% w/v). Liver was perfused *in situ* through the portal vein with two different perfusion media: T1 solution pH 7.4 (0.9% NaCl, 0.05% KCl, 0.2% HEPES, 0.08 mg ml<sup>-1</sup> EGTA) and T2 solution pH 7.4 (0.6% NaCl, 0.05% KCl, 1.2% HEPES, 0.07% CaCl<sub>2</sub>, 3 g ml<sup>-1</sup> collagenase type I). Flow rate was 5 ml min<sup>-1</sup>. After collagenase digestion, hepatocytes were filtered through a 70 µm cell strainer and passed on a 37.5% PERCOLL cushion (Amersham). Pellet of viable cells was resuspended in DMEM and recovered at 5% CO<sub>2</sub>, 37 °C for 2 h before starting the experimental procedure.

**Analysis of ribosomal biogenesis.** Processing of rRNA was analysed by a pulse-chase experiment with <sup>3</sup>H-uridine (Amersham) with a modified method. Exponentially growing MEFs were labelled with 3 µCi ml<sup>-1</sup> <sup>3</sup>H-uridine and chased for the indicated period. Total RNA was extracted by TRIzol (Sigma) according to the manufacturer's protocol. The same counts per minute were loaded on a 1% formaldehyde-agarose gel and then transferred onto a nylon membrane HYBOND-N (Amersham) for 16 h. RNA was crosslinked by ultraviolet irradiation and nylon membrane was treated with EN3HANCE spray (Perkin Elmer) before autoradiography.

**S-phase entry analysis by the <sup>3</sup>H-thymidine assay.** The assay was performed at early passages (P1–P2), and <sup>3</sup>H-thymidine uptake was detected. MEFs were seeded at 60% confluency in 6-well plate and starved for 48 h in low serum. Synchronized cells were stimulated with 10% serum, and 1 µCi ml<sup>-1</sup> of <sup>3</sup>H-thymidine (Amersham) was added for 1 h at 6, 9, 12, 15, 18, 21 and 24 h. After 1 h, cells were incubated for 15 min with 5% trichloroacetic acid to precipitate DNA onto the dishes, and after two washes in 70% cold-ethanol macromolecules were solubilized in 1 ml solubilization buffer (0.1 M NaOH, 2% Na<sub>2</sub>CO<sub>3</sub>, 1% SDS) for 1 h at 60 °C. The mixture was neutralized with 36% of neutralizing solution (1 M HCl) and half of each radioactive sample was solubilized in scintillation fluid and counted with a  $\beta$ -counter. Data were derived from the mean of three independent experiments and normalized to the total amount of DNA.

28. Colombo, E., Marine, J. C., Danovi, D., Falini, B. & Pelicci, P. G. Nucleophosmin regulates the stability and transcriptional activity of p53. *Nature Cell Biol.* **4**, 529–533 (2002).
29. Betts, D. H., Barcroft, L. C. & Watson, A. J. Na/K-ATPase-mediated 86Rb<sup>+</sup> uptake and asymmetrical trophectoderm localization of  $\alpha$ 1 and  $\alpha$ 3 Na/K-ATPase isoforms during bovine preattachment development. *Dev. Biol.* **197**, 77–92 (1998).
30. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* **88**, 593–602 (1997).
31. De Palma, M. & Naldini, L. Transduction of a gene expression cassette using advanced generation lentiviral vectors. *Methods Enzymol.* **346**, 514–529 (2002).
32. Ceci, M. *et al.* Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly. *Nature* **426**, 579–584 (2003).
33. Rim, J. S., Mynatt, R. L. & Gawronska-Kozak, B. Mesenchymal stem cells from the outer ear: a novel adult stem cell model system for the study of adipogenesis. *FASEB J.* **19**, 1205–1207 (2005).
34. Strezoska, Z., Pestov, D. G. & Lau, L. F. Functional inactivation of the mouse nucleolar protein Bop1 inhibits multiple steps in pre-rRNA processing and blocks cell cycle progression. *J. Biol. Chem.* **277**, 29617–29625 (2002).

# CDK targets Sae2 to control DNA-end resection and homologous recombination

Pablo Huertas<sup>1</sup>, Felipe Cortés-Ledesma<sup>2</sup>, Alessandro A. Sartori<sup>1†</sup>, Andrés Aguilera<sup>2</sup> & Stephen P. Jackson<sup>1</sup>

DNA double-strand breaks (DSBs) are repaired by two principal mechanisms: non-homologous end-joining (NHEJ) and homologous recombination (HR)<sup>1</sup>. HR is the most accurate DSB repair mechanism but is generally restricted to the S and G2 phases of the cell cycle, when DNA has been replicated and a sister chromatid is available as a repair template<sup>2–5</sup>. By contrast, NHEJ operates throughout the cell cycle but assumes most importance in G1 (refs 4, 6). The choice between repair pathways is governed by cyclin-dependent protein kinases (CDKs)<sup>2,3,5,7</sup>, with a major site of control being at the level of DSB resection, an event that is necessary for HR but not NHEJ, and which takes place most effectively in S and G2 (refs 2, 5). Here we establish that cell-cycle control of DSB resection in *Saccharomyces cerevisiae* results from the phosphorylation by CDK of an evolutionarily conserved motif in the Sae2 protein. We show that mutating Ser 267 of Sae2 to a non-phosphorylatable residue causes phenotypes comparable to those of a *sae2Δ* null mutant, including hypersensitivity to camptothecin, defective sporulation, reduced hairpin-induced recombination, severely impaired DNA-end processing and faulty assembly and disassembly of HR factors. Furthermore, a Sae2 mutation that mimics constitutive Ser 267 phosphorylation complements these phenotypes and overcomes the necessity of CDK activity for DSB resection. The Sae2 mutations also cause cell-cycle-stage specific hypersensitivity to DNA damage and affect the balance between HR and NHEJ. These findings therefore provide a mechanistic basis for cell-cycle control of DSB repair and highlight the importance of regulating DSB resection.

To initiate HR, one strand of the broken DNA duplex is resected in the 5'→3' direction, generating single-stranded DNA (ssDNA) that can anneal with a homologous DNA duplex<sup>8</sup>. In *S. cerevisiae*, effective resection and HR require sustained Cdc28/Clb (Cdk1/cyclin B) kinase activity<sup>2,3,5</sup>, although the CDK targets mediating this control are still unknown. One potential target is Sae2, a protein first identified as being required for meiotic recombination. Sae2 controls the initiation of DNA-end resection in meiotic and mitotic cells<sup>9–12</sup> and was recently shown to be a DNA endonuclease<sup>13</sup>. Previous work has shown that Sae2 is targeted by the Mec1 and Tel1 kinases in response to DNA damage, generating forms of Sae2 with decreased mobility in SDS-polyacrylamide gels<sup>14</sup>. Such alterations in Sae2 gel-mobility also occurred in unperturbed cycling cells, specifically in S and G2, indicating that Sae2 might be a Cdc28 target (data not shown, and Supplementary Fig. 2a). In accord with this idea, the amount of slower-migrating Sae2 was diminished when Cdc28 was inactivated in G2-synchronized cultures by galactose-driven expression of the Cdc28/Clb repressor, Sic1 (ref. 15) (Fig. 1a).

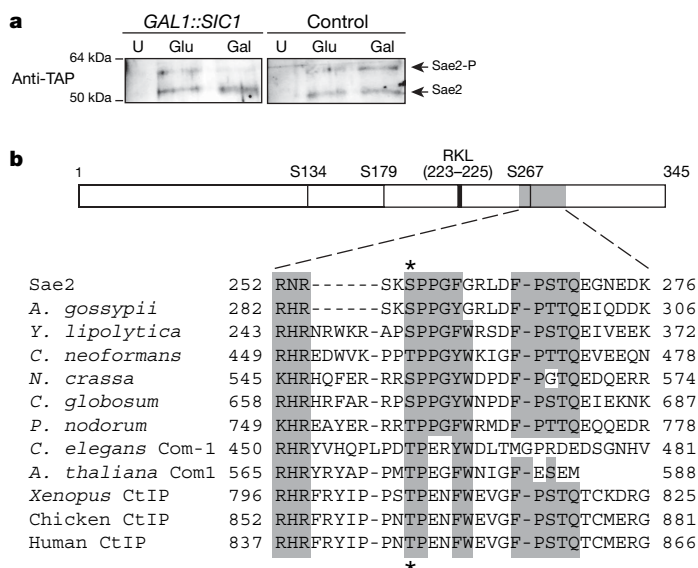
Sae2 contains three potential CDK phosphorylation sites (Fig. 1b and Supplementary Fig. 1); two of these—Ser 267 and Ser 134—received the highest scores for predicted phosphorylation sites in the protein (Supplementary Table 1). Ser 267 maps to the Sae2 region

most highly conserved with its non-yeast orthologues (Fig. 1b), which include human CtIP, *Caenorhabditis elegans* Com1 and *Arabidopsis thaliana* Com1 (refs 16–18). To address the possible function(s) of Ser 267 and other potential target sites for CDK in Sae2, we generated yeast strains in which each site was individually mutated to a non-phosphorylatable alanine residue. The *sae2-S267A* mutant showed strong hypersensitivity towards the topoisomerase I inhibitor camptothecin (Fig. 1c and Supplementary Fig. 2b; Supplementary Fig. 2c shows that this mutant is nearly as sensitive as the *sae2Δ* strain). By contrast, *sae2-S134A*, *sae2-S179A* and *sae2-S134A,S179A* cells did not show detectable hypersensitivity to camptothecin, and combining *sae2-S267A* with these other mutations showed no synergistic effect (Supplementary Fig. 2b). When we mutated Ser 267 to glutamic acid to mimic constitutive phosphorylation, the resulting strain displayed no detectable hypersensitivity to camptothecin (Fig. 1c and Supplementary Fig. 2c). Together with the fact that mutation of Ser 267 did not alter Sae2 protein expression (Fig. 1d), these data suggested that Ser 267 phosphorylation is required for Sae2 function. Furthermore, as the major cytotoxic lesions for camptothecin are DSBs arising when replication forks encounter trapped topoisomerase I–DNA complexes<sup>19</sup>, these results suggested that phosphorylation of Sae2 on Ser 267 is important for responses to DSBs generated during S phase. Indeed, the *sae2*-null and *sae2-S267A* strains, but not the *sae2-S267E* strain, showed hypersensitivity to methyl methanesulphonate (MMS), which also yields DSBs in S phase (Supplementary Fig. 2d). Consistent with analogous residues controlling the activity of Sae2-related proteins in other species, human U2OS cells downregulated for endogenous CtIP and expressing short interfering RNA (siRNA)-resistant GFP–CtIP–T847A were as sensitive to camptothecin as control cells expressing GFP (green fluorescent protein) alone, whereas cells expressing a phospho-mimicking CtIP derivative (GFP–CtIP–T847E) showed higher resistance to camptothecin (Fig. 1e; for expression levels and downregulation see Supplementary Fig. 2e).

Efficient phosphorylation of CDK substrates *in vivo* often requires the binding of cyclin to an Arg-X-Leu (RXL) motif in the target<sup>20</sup>. Such a motif is present upstream of Ser 267 in Sae2 (Fig. 1b and Supplementary Fig. 1), and mutating this motif (*Sae2-R223A,L225A*) caused hypersensitivity to camptothecin as severe as that of *sae2-S267A* or *sae2Δ* cells (Fig. 1c) even though the mutated proteins were expressed at normal levels (Fig. 1d). Furthermore, the camptothecin hypersensitivity caused by the *Sae2-R223A,L225A* mutation was largely suppressed when the protein also contained the phospho-mimicking Sae2 S267E mutation (Fig. 1c). Collectively, these findings strongly suggested that Sae2 function requires its modification by CDK–cyclin complexes.

To examine the phosphorylation of Sae2 on Ser 267 directly, we raised a phosphospecific antibody against this site (γS267). Western

<sup>1</sup>The Wellcome Trust and Cancer Research UK Gurdon Institute, and Department of Zoology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. <sup>2</sup>Centro Andaluz de Biología Molecular y Medicina Regenerativa CABIMER, Universidad de Sevilla-CSIC, Avenida Américo Vespucio s/n, 41092 Sevilla, Spain. <sup>†</sup>Present address: Institute of Molecular Cancer Research, University of Zurich-Irchel, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.



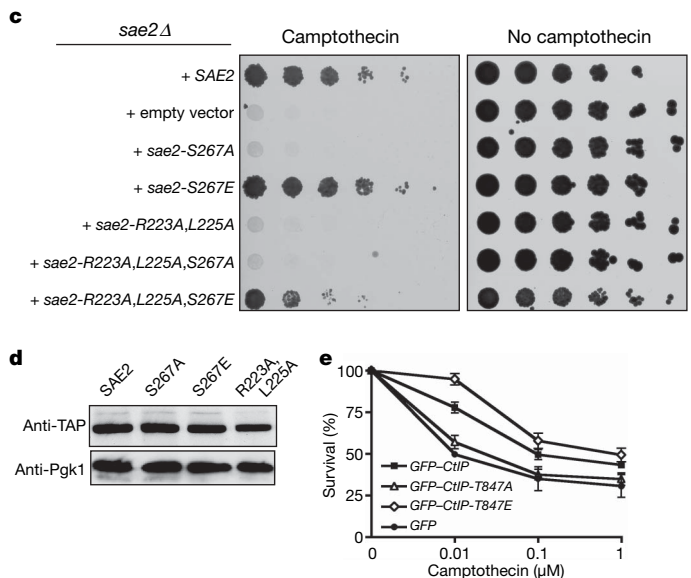
**Figure 1 | Ser 267 mutation impairs Sae2 function.** **a**, Left: TAP-tagged Sae2 was purified from cells expressing galactose-inducible *SIC1* (Gal) or not expressing *SIC1* (Glu). U, control untagged strain. Right: as above, but the strain lacked galactose-inducible *SIC1*. **b**, Sae2 diagram and homology to orthologues (see Methods for full alignment). *S. cerevisiae*, *Saccharomyces cerevisiae*; *A. gossypii*, *Ashbya gossypii*; *Y. lipolytica*, *Yarrowia lipolytica*; *C. neoformans*, *Cryptococcus neoformans*; *N. crassa*, *Neurospora crassa*; *C. globosum*, *Chaetomium globosum*; *P. nodorum*, *Phaeosphaeria nodorum*;

immunoblotting revealed that this antibody specifically detected immunoprecipitated wild-type Sae2 but not the Sae2-S267A or Sae2-R223A,L225A proteins (Fig. 2a). Furthermore, the antibody detected the slower-migrating form of Sae2 that was present at elevated levels in G2-synchronized cultures (Fig. 2b), indicating that phosphorylation of Sae2 on Ser 267 is subject to cell-cycle control. Notably, Sae2 immunoprecipitation recovered cyclins Clb3 and Clb2 from extracts prepared from G2-synchronized cells but not from G1 cells (Fig. 2b). Also consistent with Sae2's being a direct target of CDK, incubation of purified glutathione S-transferase-fused Sae2 protein—but not the S267A mutant—with recombinant CDK–cyclin complexes and ATP produced Ser 267 phosphorylation (Fig. 2c).

Further analyses suggested that all aspects of Sae2 function require phosphorylation at Ser 267. Thus, like *sae2Δ* cells, *sae2-S267A* cells were severely compromised in hairpin-induced recombination<sup>21</sup>, whereas *sae2-S267E* cells behaved similarly to the wild type (Fig. 2d). Furthermore, whereas homozygous diploid *sae2-S267E* cells produced viable spores at levels similar to those produced by the wild-type strain, like *sae2Δ* cells<sup>10,12</sup>, the *sae2-S267A* homozygous mutant strain showed a severe sporulation defect and almost no spore viability (Figs 2e and 2f). A defect in spore viability due to mutations in Ser 267 or Pro 268 has also been reported recently<sup>18</sup>.

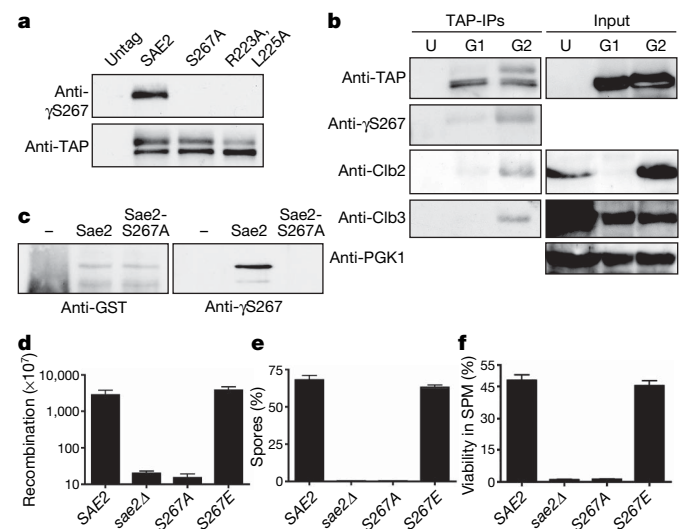
Sae2 regulates resection of chromosomal DSBs formed by the HO endonuclease<sup>9</sup>. To determine whether Sae2 Ser 267 controls this function, we generated an irreparable HO-induced DSB in the *MAT* locus and analysed resulting samples with a neutral dot-blot approach that detected only resected DNA<sup>22</sup>. For this we used three probes: one directly adjacent to the HO cleavage site, one 5 kilobases (kb) downstream and one at the *LEU2* locus, about 100 kb distal from the HO site, as a negative control. Significantly, *sae2Δ* cells and cells bearing the *sae2-S267A* mutation were impaired in resection close to the HO site, and this impairment was even more pronounced when assayed 5 kb away (Fig. 3a). By contrast, the *sae2-S267E* mutant carried out resection almost as efficiently as the wild-type strain (Fig. 3a and Supplementary Fig. 3a).

To test whether Cdc28/Cdk1-mediated Sae2 phosphorylation promotes DSB resection, we assessed resection in a strain expressing a



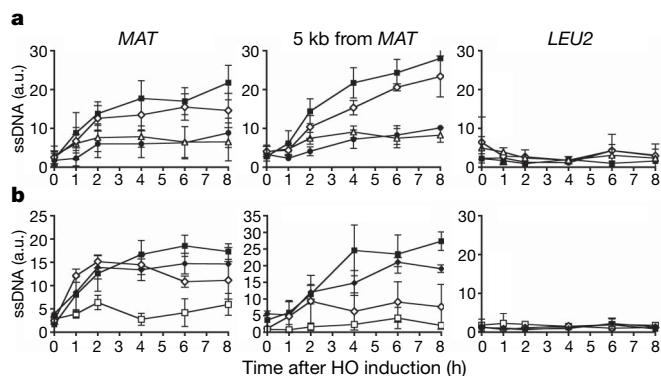
*C. elegans*, *Caenorhabditis elegans*; *A. thaliana*, *Arabidopsis thaliana*; *Xenopus*, *Xenopus laevis*; chicken, *Gallus gallus*; human, *Homo sapiens*. **c**, Fivefold serial dilutions of *sae2Δ* cultures containing the indicated SAE2 variants plated on medium lacking or containing camptothecin (5 μg ml<sup>-1</sup>). **d**, Extracts of cells harbouring TAP-tagged Sae2 variants were western immunoblotted as indicated. **e**, Survival of U2OS cells expressing GFP–CtIP fusions to 1 h treatments with the indicated doses of camptothecin. Error bars indicate s.d. (*n* = 2).

Cdc28 derivative (*cdc28-as1*) that can be specifically inhibited by the ATP analogue 1NM-PP1 (ref. 23). Whereas inhibition of Cdk1 markedly curtailed end resection in a strain expressing wild-type Sae2, it had little effect in the *sae2-S267E* mutant when resection was measured close to the HO site (Fig. 3b and Supplementary Fig. 3b). Nevertheless, the inhibitor still had some effect on the *sae2-S267E* strain when resection was assessed at the 5 kb distal site (Fig. 3b and



**Figure 2 | Sae2 is phosphorylated by Cdc28 on Ser 267.** **a**, TAP-tagged Sae2 derivatives were immunoprecipitated and detected as indicated. **b**, TAP-tagged Sae2 was purified from G1 or G2 arrested cultures. U, G2 arrested untagged control cells. Immunoprecipitated samples and inputs (5%) were immunoblotted as indicated. **c**, Glutathione S-transferase (GST)-fused Sae2 and Sae2-S267A were purified, incubated with recombinant Cdk2/Cyclin A and ATP, resolved by 10% SDS–PAGE and immunoblotted as indicated. **d**, Recombination frequencies of strains in a hairpin-containing recombination system<sup>22</sup>. **e**, Spores after 24 h in sporulation medium. **f**, Spore viability 24 h after the addition of sporulation medium (SPM)<sup>10</sup>. Error bars in **d–f** represent s.d. (*n* = 2).





**Figure 3 | DNA-end resection is controlled by Sae2.** **a**, Resection-mediated ssDNA formation at an HO DSB in wild-type SAE2 (filled squares), *sae2*-S267A (open triangle), *sae2*-S267E (open diamonds) or empty vector (solid circles) at indicated times after HO induction at the *MAT* locus (left), 5 kb downstream of *MAT* (centre) or *LEU2* locus (right). **b**, Wild-type SAE2 (squares) or *sae2*-S267E (diamonds) strains containing Cdc28-as1 were grown as in **a** but in the presence of dimethylsulphoxide (filled symbols) or 1NM-PP1 (open symbols). Results are shown as means  $\pm$  s.d. ( $n = 5$ ).

Supplementary Fig. 3b). We therefore conclude that phosphorylation of Sae2 on Ser 267 by Cdc28/Cdk1 is required for effective DSB resection but that additional Cdk1 target sites are required for resection to take place optimally. One candidate for such an additional CDK target is Rad9, which is phosphorylated by Cdc28 and was recently shown to affect DSB resection<sup>24</sup>.

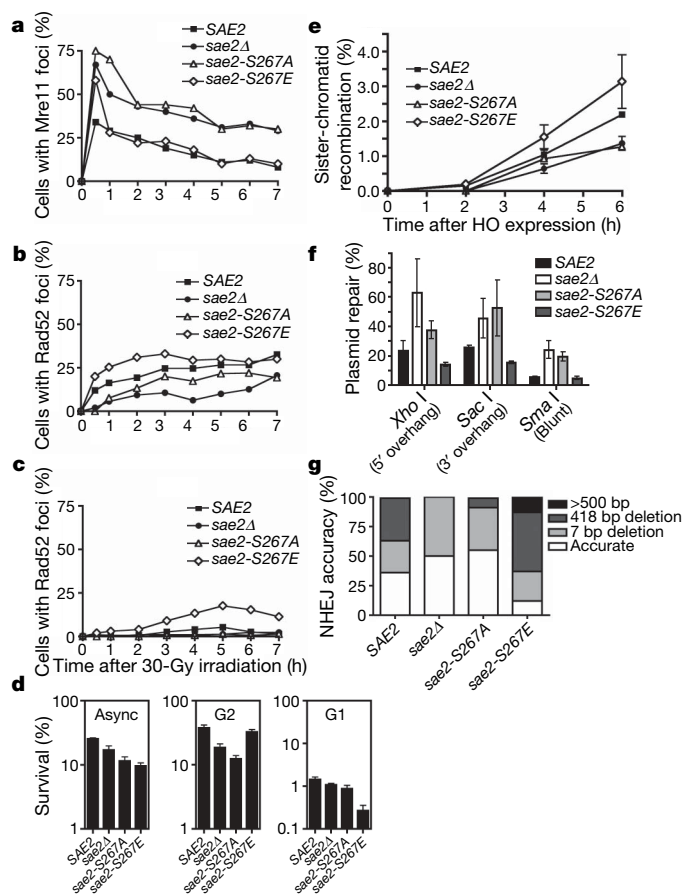
*S. cerevisiae* Mre11 is recruited quickly to DSB sites, then replaced by the HR protein Rad52 as ssDNA is formed in S and G2 cells<sup>25</sup>. We found that, like SAE2 deletion<sup>25</sup>, the *sae2*-S267A mutation caused Mre11 foci to persist longer than in wild-type cells (Fig. 4a) and delayed Rad52 focus formation in S/G2 after X-ray treatment (Fig. 4b). In contrast, the *sae2*-S267E mutant displayed Mre11 focus disassembly kinetics similar to the wild-type strain (Fig. 4a), and in fact reproducibly formed Rad52 foci faster in S and G2 than the wild-type strain (Fig. 4b). As expected, essentially no Rad52 foci were detected in wild-type G1 cells, because DSBs are not efficiently resected. By contrast—and unlike the *sae2*-S267A mutant—the *sae2*-S267E mutant formed Rad52 foci in G1 at later time points (Fig. 4c), confirming that DNA ends are processed to some degree in this mutant even in the absence of active CDK (Fig. 3b).

To address the impact of aberrant DSB processing on DSB repair, we irradiated asynchronous, G2-arrested and G1-arrested cell cultures, then kept cells in the same state (asynchronous, G2 or G1) for 6 h to allow DNA repair. Cells were subsequently plated and colony formation was used to determine survival. When we analysed asynchronous cultures, the *sae2*Δ strain showed moderate hypersensitivity to radiation (as shown previously<sup>26</sup>), and similar, although slightly more pronounced, hypersensitivities were displayed by the *sae2*-S267A and *sae2*-S267E strains (Fig. 4d). Moreover, the two Ser 267 mutations had markedly different effects in G1 and G2 (Fig. 4d). Thus, whereas the *sae2*-S267A strain was more sensitive to radiation than the control strain in G2, little or no G2 hypersensitivity was shown by the *sae2*-S267E strain. In contrast, *sae2*-S267E cells showed marked hypersensitivity to radiation in G1, whereas the *sae2*-null and *sae2*-S267A cells did not.

The above data suggested that the *sae2*-S267A and *sae2*-null strains, but not the *sae2*-S267E strain, are defective in HR because of impaired DSB resection in G2, and also suggested that the hypersensitivity of the *sae2*-S267E mutant to radiation in G1 reflects aberrant DSB resection, thus impairing NHEJ and/or triggering futile attempts to carry out HR in the absence of a sister chromatid. To test these ideas, we determined HR and NHEJ efficiencies in various *sae2* mutant backgrounds. When we used an assay in which HR intermediates were monitored by Southern blot analysis<sup>27</sup>, the *sae2*-null and *sae2*-S267A mutants showed delayed HR, whereas the *sae2*-S267E mutant showed slightly accelerated recombination (Fig. 4e and Supplementary Fig. 4).

Furthermore, by measuring NHEJ with an *in vivo* plasmid-recircularization assay<sup>28</sup>, we found that cells lacking Sae2 or bearing the S267A mutation had enhanced ( $P < 0.05$  compared with wild type) NHEJ efficiencies, regardless of whether the DSB contained a 5' overhang, a 3' overhang or a blunt end (Fig. 4f). In addition, and in agreement with an increased propensity for DSB resection that would impair NHEJ, *sae2*-S267E mutant cells showed a decrease in NHEJ efficiency when overhang substrates were used ( $P < 0.05$ ; Fig. 4f). By retrieving repaired plasmids from independent clones and sequencing them, we found that in wild-type cells most repair took place accurately or by micro-homology-mediated end-joining involving pairs of 4-bp repeats separated by 7 or 418 bp to create small or moderate deletions (Fig. 4g). In agreement with less efficient resection taking place in the *sae2*-null and *sae2*-S267A mutant strains, such cells favoured accurate repair or repair involving small (7-bp) deletions ( $P < 10^{-8}$ ; Fig. 4g; similar data were reported for *sae2*Δ cells<sup>29</sup>). In contrast, *sae2*-S267E mutant cells showed little accurate NHEJ and, instead, most repair products contained larger deletions of up to 2 kb ( $P < 10^{-4}$ ; Fig. 4g).

Thus, Cdc28/Cdk1-mediated Sae2 phosphorylation modulates the balance between NHEJ and HR during the cell cycle. These results lend strong support to models in which the commitment to DSB resection is highly regulated to ensure that the cell engages the most



**Figure 4 | Sae2 mutations affect Mre11 and Rad52 dynamics, and DSB repair.** **a**, **b**, Percentages of S/G2 cells containing Mre11 (**a**) or Rad52 (**b**) foci. **c**, Percentage of G1 cells containing a Rad52 focus. **d**, Survival of *sae2*Δ mutants containing wild-type SAE2, *sae2*-S267A, *sae2*-S267E or empty vector grown asynchronously (Async) or arrested in G1 or G2, after irradiation with 300 Gy. Error bars represent s.d. ( $n = 2$ ). **e**, Sister-chromatid recombination measured as described previously<sup>27</sup>. Standard deviations of two independent experiments are shown (see Supplementary Fig. 4 for details and representative blot). **f**, Plasmid cleaved by *Xho*I, *Sac*I or *Sma*I was transformed into strains and NHEJ efficiency was measured. Means and s.d. of three independent experiments are shown. **g**, Classes of plasmid rejoining products from 50 independent clones of each strain transformed with the *Xho*I-cut plasmid.

appropriate DNA repair pathway, thereby optimizing genome stability. As Sae2 has endonuclease activity<sup>13</sup>, we favour a model in which Sae2, possibly in cooperation with the Mre11–Rad50–Xrs2 (MRX) complex, facilitates resection in S/G2 by mediating an endonucleolytic cleavage close to the DNA break, thus generating a clean end that can serve as an efficient substrate for nucleases such as MRX and Exo1. Sae2 activity might be particularly important to initiate resection at DSBs that contain covalently bound proteins that would otherwise resist exonuclease action; indeed, this would explain why deletion of *SAE2* causes defective removal of Spo11–DNA adducts during meiosis and marked hypersensitivity to camptothecin. Sae2 might also initiate resection at radiation-induced DSBs that are resistant to exonucleases because they bear protein–DNA crosslinks or complex damage to bases at their termini. By contrast, at sites of clean DSBs, *SAE2* deletion would only slow down resection and ensuing HR, thus explaining why *sae2* mutants are not as sensitive to radiation as other HR mutants<sup>26</sup>. Finally, we note that the motif encompassing Ser 267 of Sae2 is highly conserved in Sae2 counterparts in higher eukaryotes, and that mutation of the analogous Thr 847 site in human CtIP to Ala (but not to Glu) yields hypersensitivity to camptothecin. This suggests that analogous CDK-control mechanisms for DSB resection operate in many other organisms. One exception to this, however, is likely to be provided by *Schizosaccharomyces pombe*, whose Sae2/CtIP homologue, Ctp1, lacks a CDK site analogous to Ser 267 of Sae2. In this case, it seems that, rather than controlling Ctp1 phosphorylation, the CDK machinery instead regulates the protein expression of Ctp1 (ref. 30). Nevertheless, although some species-specific variations undoubtedly exist, we speculate that Sae2/CtIP/Com1/Ctp1 proteins will turn out to have ubiquitous functions in facilitating DSB resection in S and G2 and modulating the choice of DSB repair pathway in eukaryotic cells.

## METHODS SUMMARY

A *sae2Δ* strain in W303 background<sup>9,14</sup> was transformed with plasmids harbouring the indicated *SAE2* mutant and used in all experiments except those listed below. For Figs 1a and 2b, a Sae2-TAP strain (Open Biosystems) was used. In Fig. 2e, f, a strain harbouring the indicated Sae2 mutant at its chromosomal locus in the SK1 background was used. For Fig. 3, we deleted *SAE2* in a strain harbouring the *cdc28as1* allele in the JKM179 background<sup>5</sup>. The W5573-15D strain was used in Fig. 4a–c (ref. 25). A *sae2*-deleted OIS-15 strain was used in Fig. 4e (ref. 27). Yeasts were grown with standard procedures. When indicated, cells were arrested in G1 with  $\alpha$ -factor and in G2 with nocodazole. DNA resection assays<sup>22</sup>, focus formation<sup>25</sup>, recombination between sister chromatids<sup>27</sup> and NHEJ assays<sup>30</sup> were as described previously. CtIP downregulation was as previously reported<sup>17</sup>. Western blotting was by standard methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 21 February; accepted 26 June 2008.

Published online 20 August 2008.

- Shrivastav, M., De Haro, L. P. & Nickoloff, J. A. Regulation of DNA double-strand break repair pathway choice. *Cell Res.* **18**, 134–147 (2008).
- Aylon, Y., Liefshitz, B. & Kupiec, M. The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. *EMBO J.* **23**, 4868–4875 (2004).
- Caspari, T., Murray, J. M. & Carr, A. M. Cdc2-cyclin B kinase activity links Crb2 and Rqh1-topoisomerase III. *Genes Dev.* **16**, 1195–1208 (2002).
- Hinz, J. M., Yamada, N. A., Salazar, E. P., Tebb, R. S. & Thompson, L. H. Influence of double-strand-break repair pathways on radiosensitivity throughout the cell cycle in CHO cells. *DNA Repair (Amst.)* **4**, 782–792 (2005).
- Ira, G. *et al.* DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. *Nature* **431**, 1011–1017 (2004).
- Karathanasis, E. & Wilson, T. E. Enhancement of *Saccharomyces cerevisiae* end-joining efficiency by cell growth stage but not by impairment of recombination. *Genetics* **161**, 1015–1027 (2002).
- Esashi, F. *et al.* CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. *Nature* **434**, 598–604 (2005).
- Aylon, Y. & Kupiec, M. DSB repair: the yeast paradigm. *DNA Repair (Amst.)* **3**, 797–815 (2004).

- Clerici, M., Mantiero, D., Lucchini, G. & Longhese, M. P. The *Saccharomyces cerevisiae* Sae2 protein promotes resection and bridging of double strand break ends. *J. Biol. Chem.* **280**, 38631–38638 (2005).
- McKee, A. H. & Kleckner, N. A general method for identifying recessive diploid-specific mutations in *Saccharomyces cerevisiae*, its application to the isolation of mutants blocked at intermediate stages of meiotic prophase and characterization of a new gene *SAE2*. *Genetics* **146**, 797–816 (1997).
- Neale, M. J., Pan, J. & Keeney, S. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053–1057 (2005).
- Prinz, S., Amon, A. & Klein, F. Isolation of *COM1*, a new gene required to complete meiotic double-strand break-induced recombination in *Saccharomyces cerevisiae*. *Genetics* **146**, 781–795 (1997).
- Lengsfeld, B. M., Rattray, A. J., Bhaskara, V., Ghirlando, R. & Paull, T. T. Sae2 is an endonuclease that processes hairpin DNA cooperatively with the Mre11/Rad50/Xrs2 complex. *Mol. Cell* **28**, 638–651 (2007).
- Baroni, E., Viscardi, V., Cartagena-Lirola, H., Lucchini, G. & Longhese, M. P. The functions of budding yeast Sae2 in the DNA damage response require Mec1- and Tel1-dependent phosphorylation. *Mol. Cell. Biol.* **24**, 4151–4165 (2004).
- Mendenhall, M. D. & Hodge, A. E. Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **62**, 1191–1243 (1998).
- Penkner, A. *et al.* A conserved function for a *Caenorhabditis elegans* Com1/Sae2/CtIP protein homolog in meiotic recombination. *EMBO J.* **26**, 5071–5082 (2007).
- Sartori, A. A. *et al.* Human CtIP promotes DNA end resection. *Nature* **450**, 509–514 (2007).
- Uanschou, C. *et al.* A novel plant gene essential for meiosis is related to the human CtIP and the yeast *COM1/SAE2* gene. *EMBO J.* **26**, 5061–5070 (2007).
- Pommier, Y. Topoisomerase I inhibitors: camptothecins and beyond. *Nature Rev. Cancer* **6**, 789–802 (2006).
- Chen, J., Saha, P., Kornbluth, S., Dynlacht, B. D. & Dutta, A. Cyclin-binding motifs are essential for the function of p21<sup>CIP1</sup>. *Mol. Cell. Biol.* **16**, 4673–4682 (1996).
- Lobachev, K. S., Gordenin, D. A. & Resnick, M. A. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* **108**, 183–193 (2002).
- Sugawara, N. & Haber, J. E. Repair of DNA double strand breaks: *in vivo* biochemistry. *Methods Enzymol.* **408**, 416–429 (2006).
- Bishop, A. C. *et al.* A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature* **407**, 395–401 (2000).
- Lazzaro, F. *et al.* Histone methyltransferase Dot1 and Rad9 inhibit single-stranded DNA accumulation at DSBs and uncapped telomeres. *EMBO J.* **27**, 1502–1512 (2008).
- Lisby, M., Barlow, J. H., Burgess, R. C. & Rothstein, R. Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell* **118**, 699–713 (2004).
- Rattray, A. J., McGill, C. B., Shafer, B. K. & Strathern, J. N. Fidelity of mitotic double-strand-break repair in *Saccharomyces cerevisiae*: a role for *SAE2/COM1*. *Genetics* **158**, 109–122 (2001).
- Cortes-Ledesma, F. & Aguilera, A. Double-strand breaks arising by replication through a nick are repaired by cohesin-dependent sister-chromatid exchange. *EMBO Rep.* **7**, 919–926 (2006).
- Boulton, S. J. & Jackson, S. P. *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.* **15**, 5093–5103 (1996).
- Lee, K. & Lee, S. E. *Saccharomyces cerevisiae* Sae2- and Tel1-dependent single-strand DNA formation at DNA break promotes microhomology-mediated end joining. *Genetics* **176**, 2003–2014 (2007).
- Limbo, O. *et al.* Ctp1 is a cell-cycle-regulated protein that functions with Mre11 complex to control double-strand break repair by homologous recombination. *Mol. Cell* **28**, 134–146 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. P. Longhese, R. Rothstein, K. Lobachev, M. Lichten and M. Foiani for providing strains, and R. Driscoll, S. Gravel, K. Dry and K. Miller for helpful discussions and comments on the manuscript. P.H. is the recipient of a Long-Term EMBO Fellowship. A.A.S. is supported by a Swiss National Foundation Grant. The S.P.J. laboratory is supported by grants from Cancer Research UK and the European Community (Integrated Project DNA repair, grant LSHG-CT-2005-512113). The A.A. laboratory is supported by grants from the Spanish Ministry of Science and Education (BFU2006-05260 and CDS2007-0015) and Junta de Andalucía (CVI624).

**Author Contributions** A.A.S. identified the homology between Sae2 and CtIP, cloned *SAE2* into pGEX-4T1 and made the original *sae2*-S267A and *sae2*-S267E mutations. All the experiments shown were performed by P.H. and were conceived by P.H. and S.P.J., except those on SCR analyses that were performed by F.C.-L. and conceived by F.C.-L. and A.A. P.H. and S.P.J. wrote the paper. All authors discussed and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.P.J. ([sjackson@gurdon.cam.ac.uk](mailto:sjackson@gurdon.cam.ac.uk)).

## METHODS

**TAP-tagged Sae2 immunoprecipitation.** TAP complexes were purified by a variation of previously described methods<sup>31</sup>. Cultures (250 ml) of TAP-tagged Sae2 variants were collected by centrifugation at 4 °C and resuspended in 1 volume of 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 0.15% Nonidet P40 in the presence of protease inhibitor (Roche) and phosphatase inhibitors. Extracts were prepared with a One-Shot cell disruptor (Constant Systems) and centrifuged for 1 h at 3,000 r.p.m. (1,400g) and 4 °C. Next, samples were incubated for 2 h at 4 °C with IgG-Sepharose (Amersham) pre-equilibrated in the same buffer. The matrix was then packed in a column and washed with 50 ml of the same buffer at 4 °C. Next, the resin was resuspended in 100 µl of 10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Nonidet P40, 0.5 mM EDTA, 1 mM dithiothreitol, transferred to a microcentrifuge tube, incubated for 2 h at 16 °C and then overnight at 4 °C with 10 U of TEV protease (Qiagen) to release Sae2 complexes from the beads. Through this procedure, Sae2 retained half of the TAP tag that could then be detected with anti-TAP antibody (Open Biosystems). Samples were centrifuged for 1 min at maximum speed (2,900g) at 4 °C; the supernatant was transferred to a new tube and 100 µl of sample loading buffer was added followed by immunoblot analysis by SDS-PAGE with the following antibodies: anti-TAP, anti-γS267 (custom made; Eurogentec), anti-PGK1 (Molecular Probes), Cdc28, Clb2 and Clb3 (Santa Cruz).

**Human cell survival assays.** Human U2OS cells expressing siRNA-resistant wild-type or mutant GFP-CtIP fusions were downregulated for endogenous CtIP with a previously published siRNA<sup>17</sup>, and 72 h afterwards were exposed to doses of camptothecin for 1 h. Survivals represent the number of colonies formed after 12 days normalized with an unirradiated control.

**Sporulation efficiency.** Homozygous diploids were grown overnight in YPAD medium, washed twice with warm sporulation medium, left in sporulation medium for 24 h at 30 °C, then fixed with 50% ethanol and stained with 4,6-diamidino-2-phenylindole (DAPI). The percentage of sporulated cells was determined by microscopy<sup>10</sup>.

**DNA-end resection assay.** Cultures of *sae2Δ cdc28-as1 GAL1::HO* strain transformed with wild-type *SAE2*, *sae2-S267A*, *sae2-S267E* or empty vector were grown to mid-exponential phase in raffinose. Samples were taken at indicated times after inducing HO by the addition of galactose. DNA was isolated, of which 1 mg was blotted in neutral and denaturing conditions with a dot-blot manifold as described previously<sup>23</sup>, then hybridized with radioactively labelled probes against the *MAT* locus, 5 kb downstream of the *MAT* locus or *LEU2* locus. Signals were quantified with a FLA-5000 instrument (Fuji) and values obtained in neutral conditions were normalized to those obtained under denaturing conditions (see Supplementary Fig. 3a). When indicated, 2 h before the addition of galactose, the culture was split in two and dimethylsulphoxide or Cdc28-as1 inhibitor 1NM-PP1 (5 µM final concentration; Calbiochem) was added.

**Rad52 and Mre11 foci analyses.** *Mre11-YFP Rad52-RFP sae2Δ* strain transformed with *SAE2*, *sae2-S267A*, *sae2-S267E* or empty vector was irradiated (30 Gy) with a Faxitron (Faxitron X-ray Corporation). Samples were taken, fixed by the addition of 0.1 volume of formaldehyde, washed three times with PBS, sonicated for 10 s and mixed 1:1 with DAPI-containing mounting medium (Vector Laboratories Inc.). Microscopy was with a DeltaVision microscope (Applied Precision). A minimum of 50 G1 (unbudded) and 50 S/G2 (budded) cells were counted at each time point and for each sample.

**Survival of irradiation.** *sae2Δ* mutants transformed with wild-type *SAE2*, *sae2-S267A*, *sae2-S267E* or an empty vector were grown asynchronously or were arrested with  $\alpha$ -factor (G1) or nocodazole (G2); they were then irradiated with 300 Gy (Faxitron), kept for 6 h in the same cell-cycle stage and then plated. Colonies arising were normalized with respect to non-irradiated samples and plotted.

**NHEJ assays.** A pRS416 vector restricted with *Xho*I, *Sac*I or *Sma*I was transformed into cells harbouring various *sae2* mutations. The number of colonies formed after 3 days was normalized with the number of colonies obtained in a parallel transformation with a circular pRS416 plasmid. Plasmids from 50 independent clones of each strain transformed with a *Xho*I-restricted plasmid as described previously were isolated and sequenced.

31. Puig, O. *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218–229 (2001).



# Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease

Chun Tang<sup>1†</sup>, John M. Louis<sup>1</sup>, Annie Aniana<sup>1</sup>, Jeong-Yong Suh<sup>1</sup> & G. Marius Clore<sup>1</sup>

HIV-1 protease processes the Gag and Gag-Pol polyproteins into mature structural and functional proteins, including itself, and is therefore indispensable for viral maturation<sup>1,2</sup>. The mature protease is active only as a dimer<sup>3–5</sup> with each subunit contributing catalytic residues<sup>6</sup>. The full-length transframe region protease precursor appears to be monomeric yet undergoes maturation via intramolecular cleavage of a putative precursor dimer<sup>5,7–11</sup>, concomitant with the appearance of mature-like catalytic activity<sup>7,9</sup>. How such intramolecular cleavage can occur when the amino and carboxy termini of the mature protease are part of an intersubunit  $\beta$ -sheet located distal from the active site is unclear. Here we visualize the early events in N-terminal autoprocessing using an inactive mini-precursor with a four-residue N-terminal extension that mimics the transframe region protease precursor<sup>5,12</sup>. Using paramagnetic relaxation enhancement, a technique that is exquisitely sensitive to the presence of minor species<sup>13–16</sup>, we show that the mini-precursor forms highly transient, lowly populated (3–5%) dimeric encounter complexes that involve the mature dimer interface but occupy a wide range of subunit orientations relative to the mature dimer. Furthermore, the occupancy of the mature dimer configuration constitutes a very small fraction of the self-associated species (accounting for the very low enzymatic activity of the protease precursor), and the N-terminal extension makes transient intra- and intersubunit contacts with the substrate binding site and is therefore available for autocleavage when the correct dimer orientation is sampled within the encounter complex ensemble.

The regulation of HIV-1 protease autoprocessing is modulated by the N-terminal flanking transframe region (TFR) sequence (Fig. 1a)<sup>2</sup>. The catalytic activity of the monomeric protease precursor is approximately three orders of magnitude less than that of the mature protease dimer (which has a monomer–dimer equilibrium dissociation constant  $K_d < 10$  nM)<sup>2,5</sup>. The appearance of mature-like catalytic activity and stable dimer formation is directly correlated with a single rate-limiting step comprising intramolecular (first order) cleavage of a putative transient dimeric precursor species at the p6<sup>pol</sup>–protease (PR) junction<sup>7,9,10</sup>. Mutations within the latter that prevent cleavage lead to the production of an N-terminally extended 17-kDa protease precursor species, and cause a severe defect in Gag polyprotein processing and the complete loss of viral infectivity *in vivo*<sup>17,18</sup>. Subsequent cleavage at the C terminus of protease at the PR–reverse transcriptase (RT) junction (Fig. 1a) occurs via an intermolecular (second order) reaction catalysed by a fully active protease dimer<sup>19</sup>. Mutations within the PR–RT junction that block C-terminal cleavage do not significantly affect either enzymatic activity and dimerization of the protease *in vitro*<sup>19,20</sup> or processing of HIV-1 precursor proteins, virus maturation, viability and morphology *in vivo*<sup>20</sup>, indicating that the presence of the C-terminal reverse transcriptase sequence has negligible influence on the protease precursor<sup>19,20</sup>. Thus, only autoprocessing at the N

terminus of protease at the p6<sup>pol</sup>–PR junction is an absolute prerequisite for stable protease dimer formation, the appearance of mature catalytic activity and complete processing of viral precursors. Before cleavage at the p6<sup>pol</sup>–PR junction, intermediate precursor forms may be liberated by intramolecular cleavage at competing sites (for example, p2–NC and TFP–p6<sup>pol</sup>; see Fig. 1a) that become available for productive binding and hydrolysis<sup>11</sup>, but these precursors will show the same low catalytic activity as that of the p6<sup>pol</sup>–PR precursor<sup>9,10</sup>.

As little as a four-residue extension at the N terminus of protease, corresponding to the C-terminal residues of p6<sup>pol</sup>, in conjunction with a D25N mutation result in an effectively monomeric species<sup>5,12</sup>. Disruption of the native protease dimer by N-terminal extension is due to removal of the protons on the secondary amine of the N-terminal proline residue, disrupting the interstrand hydrogen bond between the amine of the N-terminal proline of one subunit and the C-terminal carbonyl oxygen of the second subunit<sup>6</sup>. C-terminal extension, however, does not have an impact on this interstrand hydrogen bond because the secondary amine of Pro 1 is preserved. Therefore, we made use of the mini-precursor, bearing only the N-terminal cleavage site, to visualize the early transient events involved in autoprocessing of the protease at the p6<sup>pol</sup>–PR junction that is required for the formation of a fully active, stable protease dimer.

The optimized mini-precursor protease construct <sup>SFNF</sup>PR(D25N) comprises a four-residue N-terminal extension (Ser-Phe-Asn-Phe) derived from the TFR (Fig. 1a), a D25N mutation to abolish all residual catalytic activity, and C67A and C95A mutations to remove surface cysteines (Supplementary Fig. 1a)<sup>9,10,12</sup>. The corresponding active <sup>SFNF</sup>PR(D25) mini-precursor construct undergoes autoprocessing during expression to release the mature protease (see Methods). NMR analysis of <sup>SFNF</sup>PR(D25N) shows that it is monomeric (with an upper limit of ~10% dimer from translational diffusion measurements); the secondary and tertiary structures of the mature protease are preserved with the exception of the N- and C-terminal strands which form an intersubunit four-stranded anti-parallel  $\beta$ -sheet in the mature dimer; and residues –4 to 9 and 95–99 are disordered and highly mobile (see Methods and Supplementary Fig. 1b–e).

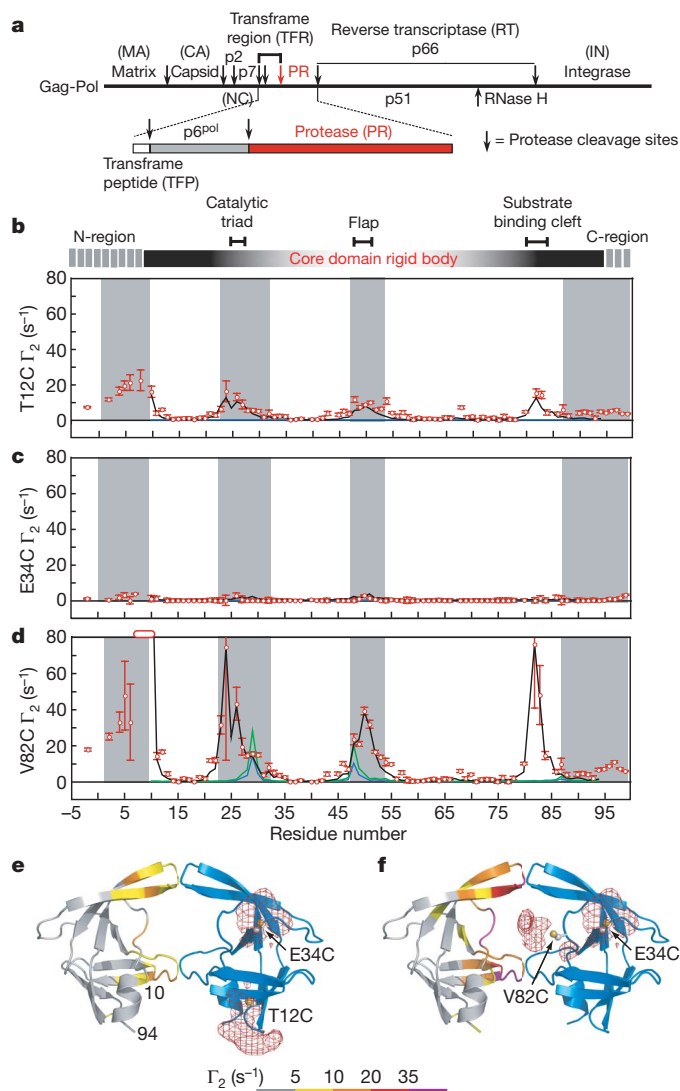
Because enzymatically active protease is dimeric, and the rate-limiting step in autoprocessing is unimolecular<sup>7,9</sup>, transient self-association of the precursor must occur to initiate autoprocessing. To visualize this phenomenon we measured intermolecular paramagnetic relaxation enhancements (PREs) by introducing a spin label via conjugation to three engineered surface-exposed cysteine residues: T12C, E34C and V82C (one at a time). These sites are frequently mutated in viable HIV-1 variants<sup>2</sup>. T12C and V82C are located at the periphery of the substrate-binding cleft in the mature dimer, whereas E34C is relatively far removed from the dimer interface (Fig. 1e, f). In a rapidly exchanging system, the PRE <sup>1</sup>H<sub>N</sub>– $\Gamma_2$  rates<sup>21</sup> are population-weighted averages of the PRE rates of the species present<sup>13,14</sup>. Because the PRE rate

<sup>1</sup>Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA.

<sup>†</sup>Present address: Department of Biochemistry, University of Missouri, Columbia, Missouri 65211, USA.

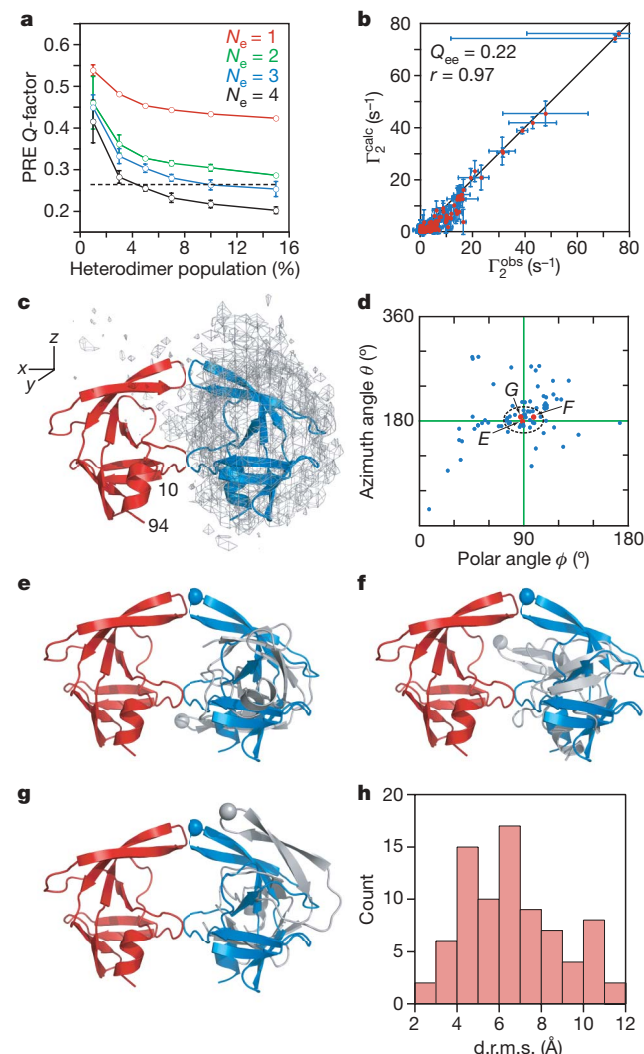
for a paramagnetic centre-proton pair is proportional to the  $\langle r^{-6} \rangle$  average of the distance between them, and the PRE effect is large owing to the high magnetic moment of an unpaired electron, the PRE in the fast exchange regime is very sensitive to the presence of lowly populated (<5%), highly transient species in solution providing there are paramagnetic centre-proton distances in the minor species that are shorter than in the predominant species<sup>13–15</sup>.

PREs were measured on a 1:1 mixture of 0.2 mM U- $[^2\text{H}/^{13}\text{C}/^{15}\text{N}]$ -labelled  $^{\text{SFNF}}$ PR(D25N) and spin-labelled  $^{\text{SFNF}}$ PR(D25N) at natural isotopic abundance. Because  $^1\text{H}_\text{N}$ - $\Gamma_2$  rates are measured using  $^1\text{H}$ - $^{15}\text{N}$  correlation-based experiments<sup>21</sup>, the observed  $^1\text{H}_\text{N}$ - $\Gamma_2$  rates arise solely from intermolecular interactions between the spin-labelled



**Figure 1 | Intermolecular PRE profiles.** **a**, Organization of the Gag-Pol polyprotein<sup>1,2</sup>. **b–d**, Intermolecular PREs observed on U- $[^2\text{H}/^{13}\text{C}/^{15}\text{N}]$ -labelled  $^{\text{SFNF}}$ PR(D25N) originating from a spin label conjugated to T12C (**b**), E34C (**c**) and V82C (**d**) of  $^{\text{SFNF}}$ PR(D25N) at natural isotopic abundance. Residues broadened beyond detection are denoted by open bars. Error bars represent 1 s.d.  $\Gamma_2$  rates back-calculated from the structure of the mature dimer (for the core residues 10–94) at populations of 1% and 2% are shown as blue and green lines, respectively. Average  $\Gamma_2$  rates derived from the top 20 structures of the  $N_e = 4$  simulated annealing calculations at a population of 5% heterodimer are shown as black lines. Grey shaded areas delineate residues that are buried at the dimer interface in the mature protease. **e**, **f**, Observed intermolecular PREs originating from the spin label attached to T12C (**e**) and V82C (**f**) colour-coded on a ribbon diagram of the mature dimer<sup>24</sup> (spin label attached to the blue subunit). Atomic probability density maps<sup>25</sup> (plotted at a threshold of 10% of maximum) showing the distribution of the spin-label oxygen radicals are shown as red meshes.

protein and the isotopically labelled protein (Fig. 1b–d). For the E34C spin label, no  $^1\text{H}_\text{N}$ - $\Gamma_2$  rates greater than  $5 \text{ s}^{-1}$  are observed (Fig. 1c); this sample therefore provides a negative control, excluding the existence of solvent PRE effects arising from diffusion and random elastic collisions, or from direct intermolecular interactions between the spin



**Figure 2 | Ensemble simulated annealing and the protease mini-precursor encounter complex ensemble.** **a**, PRE Q-factor as a function of ensemble size and population of heterodimer. Dashed line denotes the expected Q-factor when agreement between observed and calculated  $\Gamma_2$  rates is comparable to the experimental error in the measurements. **b**, Correlation between observed and calculated  $\Gamma_2$  rates for  $N_e = 4$  and a heterodimer population of 5%.  $Q_{\text{ee}}$  is the ensemble of ensembles average PRE Q-factor for the 20 calculated  $N_e = 4$  ensembles and  $r$  the correlation coefficient. Error bars in **a** and **b** represent 1 s.d. **c**, Atomic probability density map<sup>25</sup> (grey mesh, plotted at a threshold of 20% of maximum) showing the distribution of the spin-labelled subunit relative to the isotopically labelled subunit (red ribbon) in the  $^{\text{SFNF}}$ PR(D25N) encounter complexes. The location of the second subunit in the mature dimer is shown as a blue ribbon. **d**, Orientations in spherical coordinates of the vector joining the centre of masses of the two interacting molecules in the encounter complexes relative to the coordinate system shown in **c** with the z axis corresponding to the  $C_2$  symmetry axis of the mature dimer. The  $\phi, \theta$  angles for the mature dimer are located at the crosshair. **e–g**, Representative encounter complexes (labelled and denoted by red dots in **d**) corresponding to the structures with the closest spherical angles (**e**), the smallest d.r.m.s. (**f**) and the smallest atomic r.m.s. displacement (**g**) relative to the mature dimer. The C $\alpha$  atom of Gly 51 at the tip of the flap is shown as a sphere to guide the eye. The isotopically labelled and spin-labelled subunits are shown in red and grey, respectively; the blue subunit corresponds to the orientation relative to the red subunit seen in the mature dimer. **h**, Histogram of the d.r.m.s. metric for the  $N_e = 4$  structures (total of  $20 \times 4 = 80$  conformers) at a population of 5% heterodimer.

label and the U- $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labelled protein. The PRE profiles for the T12C (Fig. 1b) and V82C (Fig. 1d) spin labels are similar but the magnitude for the latter is 4- to 8-fold greater than for the former. Within the ordered core of the precursor (residues 10–94), large intermolecular PREs are observed for residues 21–30, 46–55 and 80–85 located at or close to the dimer interface. Residues 21–30 encompass the catalytic triad, residues 46–66 correspond to the flap region which gates the active site, and residues 80–81 and 83–84 are located in the substrate binding cleft (Fig. 1e, f). In addition, the N-terminal region experiences sizeable PREs from the T12C (Fig. 1b) and V82C (Fig. 1d) spin labels. These data demonstrate that transient self-association of the precursor involves residues located at the dimer interface in the mature dimer. A similar intermolecular PRE profile is observed from V82C spin-labelled, full-length TFR-PR(D25N) precursor to U- $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ - $^{\text{SFNF}}$ PR(D25N), indicating that the transient dimerization interface is preserved on further N-terminal extension of the protease precursor (Supplementary Fig. 2a).

Back-calculation of the PREs from the structure of the mature dimer shows that almost zero PRE values are expected for the T12C and E34C spin labels at a population of 1–2% mature heterodimer (Fig. 1b, c). For the V82C label, small PRE values at a population of 1–2% mature heterodimer are predicted for residues 27–30 and 48–50 (Fig. 1d, blue line). The mature dimer does not predict the large observed PRE values observed for residues 20–26, 30–35 and 80–83. Furthermore, in the mature dimer residues 80–83 of one subunit are located on the opposite side of the dimer interface from residues 80–83 of the other subunit, and thus the large intermolecular PREs observed from the V82C spin label to residues 80–83 would require a  $\sim 180^\circ$  rotation of one subunit relative to its position in the mature dimer. Thus, the upper limit of the total population of mature dimer (heterodimer and homodimer) cannot exceed 2–4%.

Transient interactions between  $^{\text{SFNF}}$ PR(D25N) precursor monomers were visualized semi-quantitatively using rigid-body simulated annealing calculations<sup>14,16,22</sup> to optimize the agreement between observed and calculated  $\Gamma_2$  rates arising from the T12C, E34C and V82C spin labels simultaneously (see Methods). The flexible N- and C-terminal regions (residues –4 to 9 and 95–99, respectively) were excluded from the calculations. A single conformer representation ( $N_c = 1$ ) for the transient dimer does not account for the PRE data and even at a heterodimer population of 15% the PRE Q-factor<sup>23</sup> (see Methods for definition) has a value of greater than 0.4 (Fig. 2a). Thus, the dimeric  $^{\text{SFNF}}$ PR(D25N) precursor is an ensemble of multiple encounter complexes. For  $N_c \geq 2$ , the average PRE Q-factor decreases rapidly as the heterodimer population is increased above 1%, levelling off at a population of  $\sim 5\%$  (Fig. 2a). The best results are obtained with  $N_c = 4$ , and larger ensemble sizes are unjustified and would result in over-fitting the data. For  $N_c = 4$ , the PRE Q-factors at a heterodimer population of 3–5% are close to the expected PRE Q-factor based on experimental error (Fig. 2a), consistent with translational diffusion data (Methods and Supplementary Fig. 1d). Given a total protein concentration of 0.4 mM, the apparent  $K_d$  for self-association is therefore 3–6 mM. A comparison of the calculated and observed PRE profiles and a correlation plot of observed versus calculated  $\Gamma_2$  rates for  $N_c = 4$  at a heterodimer population of 5% are shown in Fig. 1b–d and Fig. 2b, respectively.

The distribution of the spin-labelled monomer relative to the isotopically labelled monomer in the computed ensemble of  $^{\text{SFNF}}$ PR(D25N) encounter complexes is shown in Fig. 2c. The predominant interactions between the two monomers involve the same residues that comprise the dimer interface in the mature dimer, and one subunit of the mature dimer is embedded within the ensemble distribution of the spin-labelled subunit. The orientation of the subunits in the encounter complex ensembles can be described by spherical angles describing the orientation of the vector joining the centre of masses of the two subunits to the coordinate axis frame. Many members within the calculated ensemble are clustered around the values corresponding to the mature dimer (Fig. 2d). This is reflected in the distribution of the distance root mean square (d.r.m.s.; see Methods) deviation metric where over one-half of the

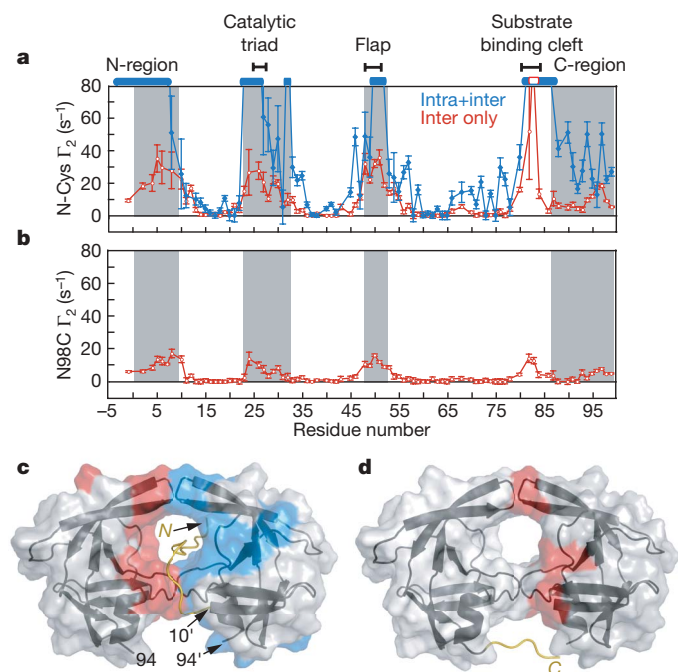
ensemble members have d.r.m.s. values less than 6 Å (Fig. 2h). However, the structures with spherical angles close to the mature dimer (indicated by arrows in Fig. 2d) and low d.r.m.s. values have a widespread range of relative self-rotations, as illustrated by three examples comprising the ensemble members with the closest spherical angles to the mature dimer (Fig. 2e), the smallest d.r.m.s. (Fig. 2f) and the smallest C $\alpha$  atomic r.m.s. displacement (Fig. 2g). The difference from the mature dimer in rotation angle about the axis joining the centre of masses of the two subunits ranges from  $13^\circ$  (Fig. 2g) to  $135^\circ$  (Fig. 2e), with an intermediate rotation angle of  $70^\circ$  for the structure in Fig. 2f (see Supplementary Fig. 3 for definitions). One can therefore conclude that the actual occupancy of a structure within the encounter complex ensemble corresponding to the mature dimer is very small.

To probe the conformational space sample by the disordered N-terminal flanking sequence of the  $^{\text{SFNF}}$ PR(D25N) precursor we introduced a spin label on a Cys residue inserted immediately after the N-terminal serine ( $^{\text{S(C)FNF}}$ PR(D25N)). PRE measurements were carried out on a 1:1 mixture of 0.2 mM U- $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ - $^{\text{SFNF}}$ PR(D25N) precursor and 0.2 mM spin-labelled, natural isotopic abundance  $^{\text{S(C)FNF}}$ PR(D25N) to detect intermolecular PREs, and on a sample of 0.2 mM spin-labelled, U- $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labelled  $^{\text{S(C)FNF}}$ PR(D25N) to observe both inter- and intramolecular PRE effects. Although the overall PRE profiles for the two samples are similar (although differences in detail are apparent), the magnitude of the PREs for the second sample is much larger than for the first, reflecting the contribution from intramolecular PREs (Fig. 3a). The N-terminal residues –4 to 9, and residues comprising the active site, flap and substrate-binding cleft, display large inter- and intramolecular PREs (Fig. 3a, c). The intermolecular PREs involving residues 82–84 are fully consistent with the large intermolecular PREs observed on the N-terminal residues from spin-labelled V82C (Fig. 1d). These data indicate that the N-terminal tail can insert itself into the active site and make transient contact with both subunits in the encounter complex ensemble. The spin label is located four residues proximal to the scissile peptide bond, and the observation that large PREs are observed for both sides of the active site (see Fig. 3c) suggests that the tail shuttles back and forth within the substrate binding cleft formed by the two subunits in the context of a dimer. Such translational movement is a functional requirement, as the protease precursor cuts the N-terminal transframe region in two major locations before cleaving its C terminus (Fig. 1a)<sup>9,10</sup>. This is confirmed by the observation of a very similar intermolecular PRE profile from full-length TFR-PR(D25N) spin-labelled at position –44, four residues downstream from the TFP-p6<sup>pol</sup> cleavage site at residues –48/–49, to U- $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ - $^{\text{SFNF}}$ PR(D25N) (Supplementary Fig. 2b).

The C-terminal region of the  $^{\text{SFNF}}$ PR(D25N) precursor was spin-labelled at N98C. The resulting intermolecular PREs are much smaller than those with the spin label at the N terminus, but the PRE profiles are similar (Fig. 3b, d). Thus, the C-terminal flexible region can also make intermolecular contacts with the active site and substrate-binding cleft in the context of the precursor encounter complex ensemble. Because the N- and C termini are highly mobile, intermolecular PREs between the N- and C termini will be significantly attenuated. Nevertheless, intermolecular PREs are observed on residues 95–97 from the spin label at the N terminus (Fig. 3a), and on residues 5–8 (Fig. 3b) from the spin label at the C terminus (Fig. 3b). Small intermolecular PREs are also observed from the N98C spin label to the C-terminal region (residues 95–99). These observations might suggest the existence of transient, loose interactions between the N- and C termini that may partially approximate a portion of the intersubunit  $\beta$ -sheet in the mature dimer.

The PRE data presented here demonstrate that although the HIV-1 protease precursor is predominantly monomeric, transient encounter complex dimers are formed using the same interface as that of the mature dimer but with a wide range of relative subunit orientations. Only a very small fraction of the encounter complexes adopt the same subunit orientation as in the mature protease, accounting for the very low enzymatic activity of the precursor. This small subset, which may be partially stabilized by transient, loose interactions involving the N- and





**Figure 3 | PRE profiles with spin labels attached at the N- and C termini of the  $^{SFNF}$ PR(D25N) mini-precursor.** **a**, Intermolecular PREs (red) observed for a 1:1 mixture (0.2 mM each) of N-terminal spin-labelled  $^{S(C)FNF}$ PR(D25N) at natural isotopic abundance and  $U\text{-}[^2H/^13C/^15N]\text{-}^{SFNF}$ PR(D25N), and the sum of the inter- and intramolecular PREs (blue) observed for 0.2 mM N-terminal spin-labelled  $U\text{-}[^2H/^13C/^15N]\text{-}^{S(C)FNF}$ PR(D25N). Residues broadened beyond detection are denoted by open bars. **b**, Intermolecular PREs observed for a 1:1 mixture (0.2 mM each) of  $U\text{-}[^2H/^13C/^15N]\text{-}^{SFNF}$ PR(D25N) and C-terminal spin-labelled (at N98C)  $^{SFNF}$ PR(D25N) at natural isotopic abundance. Grey shaded areas in **a** and **b** delineate residues that are buried at the dimer interface in the mature protease. Error bars in **a** and **b** represent 1 s.d. **c**, **d**, Inter- and intramolecular PREs with  $\Gamma_2$  rates  $>10\text{ s}^{-1}$  colour-coded in red and blue, respectively, onto the molecular surface of the mature protease dimer originating from the N-terminal (**c**) and the C-terminal (**d**) spin labels. The intramolecular PRE rates are given by the difference in PRE rates between the blue and red profiles in **a**. Cartoons of modelled N-terminal (residues -4 to 9) and C-terminal (residues 95–99) regions bearing the spin labels are included in **c** and **d**, respectively.

C-terminal regions, can accommodate transient insertion of the N-terminal region including the N-terminal cleavage site in the substrate binding cleft, thereby providing a structural model for autoprocessing at the N terminus of the protease leading to the formation of a stable dimer with mature catalytic activity.

## METHODS SUMMARY

**Sample preparation and NMR spectroscopy.** Protein expression, mutagenesis, purification and conjugation of engineered surface cysteine residues to 3-iodomethyl-(1-oxy-2,2,5,5-tetramethylpyrrolidine) are described in the Methods. Samples for NMR were in 20 mM sodium phosphate buffer, pH 5.8. NMR experiments were collected at 20 °C at a  $^1H$  spectrometer frequency of 600 MHz.  $^1H_N$  PRE data were acquired using a two-dimensional  $^1H\text{-}^{15}N$  correlation-based pulse scheme with an interleaved two time-point measurement<sup>21</sup>. **Simulated annealing calculations.** Conjoined rigid-body/torsion angle dynamics simulated annealing calculations on the basis of the PRE data were carried out using Xplor-NIH<sup>22</sup> as described<sup>14</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 7 July; accepted 12 August 2008.

1. Louis, J. M., Weber, I. T., Tozser, J., Clore, G. M. & Gronenborn, A. M. HIV-1 protease: maturation, enzyme specificity, and drug resistance. *Adv. Pharmacol.* **49**, 111–146 (2000).

2. Louis, J. M., Ishima, R., Torchia, D. A. & Weber, I. T. HIV-1 protease: structure, dynamics and inhibition. *Adv. Pharmacol.* **55**, 261–298 (2007).
3. Wlodawer, A. & Erikson, J. Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* **62**, 543–585 (1993).
4. Wlodawer, A. & Vondrasek, J. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249–284 (1998).
5. Ishima, R., Torchia, D. A., Lynch, S. M., Gronenborn, A. M. & Louis, J. M. Solution structure of the mature HIV-1 protease monomer: insight into the tertiary fold and stability of a precursor. *J. Biol. Chem.* **278**, 43311–43319 (2003).
6. Miller, M. *et al.* Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **246**, 1149–1152 (1989).
7. Louis, J. M., Nashed, N. T., Parris, K. D., Kimmel, A. R. & Jerina, D. M. Kinetics and mechanism of autoprocessing of human immunodeficiency virus type 1 protease from an analog of the Gag-Pol polyprotein. *Proc. Natl Acad. Sci. USA* **91**, 7970–7974 (1994).
8. Co, E. *et al.* Proteolytic processing mechanisms of a miniprecursor of the aspartic protease of human immunodeficiency virus type 1. *Biochemistry* **33**, 1248–1254 (1994).
9. Louis, J. M., Wondrak, E. M., Kimmel, A. R., Wingfield, P. T. & Nashed, N. T. Proteolytic processing of HIV-1 protease precursor, kinetics and mechanism. *J. Biol. Chem.* **274**, 23437–23442 (1999).
10. Louis, J. M., Clore, G. M. & Gronenborn, A. M. Autoprocessing of HIV-1 protease is tightly coupled to protein folding. *Nature Struct. Biol.* **6**, 868–874 (1999).
11. Pettit, S. C., Everitt, L. E., Choudhury, S., Dunn, B. M. & Kaplan, A. H. Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. *J. Virol.* **78**, 8477–8485 (2004).
12. Ishima, R., Torchia, D. A. & Louis, J. M. Mutational and structural studies aimed at characterizing the monomer of HIV-1 protease and its precursor. *J. Biol. Chem.* **282**, 17190–17199 (2007).
13. Iwahara, J. & Clore, G. M. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* **440**, 1227–1230 (2006).
14. Tang, C., Iwahara, J. & Clore, G. M. Visualization of transient encounter complexes in protein-protein association. *Nature* **444**, 383–386 (2006).
15. Volkov, A. N., Worall, J. A., Holtzmann, E. & Ubbink, M. Solution structure and dynamics of the complex between cytochrome c and cytochrome c peroxidase determined by paramagnetic NMR. *Proc. Natl Acad. Sci. USA* **103**, 18945–18950 (2006).
16. Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078–1082 (2007).
17. Tessmer, U. & Kräusslich, H.-G. Cleavage of human immunodeficiency virus type 1 proteinase from the N-terminally adjacent p6\* protein is essential for efficient Gag polyprotein processing and viral infectivity. *J. Virol.* **72**, 3459–3463 (1998).
18. Ludwig, C., Leiberer, A. & Wagner, G. Importance of protease cleavage sites within and flanking human immunodeficiency virus type 1 transframe protein p6\* for spatiotemporal regulation of protease activation. *J. Virol.* **82**, 4573–4584 (2008).
19. Wondrak, E. M., Nashed, N. T., Haber, M. T., Jerina, D. M. & Louis, J. M. A transient precursor of the HIV-1 protease: isolation, characterization and kinetics of maturation. *J. Biol. Chem.* **271**, 4477–4481 (1996).
20. Cherry, E. *et al.* Characterization of human immunodeficiency virus type-1 (HIV-1) particles that express protease-reverse transcriptase fusion proteins. *J. Mol. Biol.* **284**, 43–56 (1998).
21. Iwahara, J., Tang, C. & Clore, G. M. Practical aspects of  $^1H$  transverse paramagnetic relaxation enhancement measurements on macromolecules. *J. Magn. Reson.* **184**, 185–195 (2007).
22. Schwieters, C. D., Kuszewski, J. & Clore, G. M. Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **48**, 47–62 (2006).
23. Iwahara, J., Schwieters, C. D. & Clore, G. M. Ensemble approach for NMR structure refinement against  $^1H$  paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J. Am. Chem. Soc.* **126**, 5879–5896 (2004).
24. Spinelli, S., Liu, Q. Z., Alzari, P. M., Hirel, P. H. & Poljak, R. J. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. *Biochimie* **73**, 1391–1396 (1991).
25. Schwieters, C. D. & Clore, G. M. Reweighted atomic densities to represent ensembles of NMR structures. *J. Biomol. NMR* **23**, 221–225 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Ishima for providing initial backbone assignments for the  $^{SFNF}$ PR(D25N) protease construct; C. Schwieters for many discussions; Y. Sheng for help with the CS-Rosetta calculations; Y. Kim for providing the code for structure clustering and d.r.m.s. calculations; and J. Sayer for MALDI measurements. This work was supported by funds from the Intramural Program of the NIH, NIDDK and the AIDS Targeted Antiviral program of the Office of the Director of the NIH (to G.M.C.).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to G.M.C. ([mariuscc@mail.nih.gov](mailto:mariuscc@mail.nih.gov)).

## METHODS

**Vector construction and protein sample preparation for NMR studies.** Mutations T12C, E34C, V82C and N98C within the 99-amino-acid-long HIV-1 protease sequence and the mutation to insert a Cys in the flanking SFNF sequence (C-terminal residues of p6<sup>pol</sup> within the transframe region; see Fig. 1a and Supplementary Fig. 1) to generate <sup>S(C)FNF</sup>PR(D25N) were introduced in the <sup>SFNF</sup>PR(D25N) template<sup>5</sup> using the appropriate forward and reverse primers and the QuikChange kit and protocol (Stratagene). The <sup>S(C)FNF</sup>PR(D25N) construct was used because we were unable to obtain efficient spin-labelling of a precursor protein bearing an N-terminal cysteine. Mutations A(−44)C (fifth residue of p6<sup>pol</sup>) and V82C (in the protease sequence) were also introduced in the full-length TFR–PR(D25N) construct (that is, TFP–p6<sup>pol</sup>–PR(D25N); see Fig. 1a and Supplementary Fig. 1a) using the same protocol. (The TFR is 56 residues in length and adopts a random coil conformation.) The newly introduced mutations were verified both by DNA sequencing and mass spectrometry. (Note that the <sup>SFNF</sup>PR(D25N) template, in addition to the D25N mutation which eliminates all traces of catalytic activity, and the C67A and C95A mutations which remove all additional surface cysteine residues other than that to which the spin label is going to be attached, also contains three other mutations, Q7K, L33I and L63I; the latter three mutations restrict autoproteolysis of the mature protease dimer, and have been shown to have indiscernible effects on structure, stability and catalytic activity of the mature dimer<sup>10</sup>.)

*Escherichia coli* BL21 (DE3) host cells bearing the appropriate vector were grown in Luria-Bertani medium or in D<sub>2</sub>O-based minimal medium containing <sup>15</sup>N-NH<sub>4</sub>Cl and <sup>13</sup>C<sub>6</sub>,<sup>2</sup>H<sub>7</sub>-glucose as the sole nitrogen and carbon sources, respectively, at 37 °C, and induced for expression. Proteins were purified from inclusion bodies using an established protocol as described previously involving size-exclusion chromatography under denaturing conditions followed by reverse-phase HPLC<sup>9,26</sup>. Peak fractions (~0.5 mg ml<sup>−1</sup>) were stored in aliquots at −70 °C. Alternatively, two aliquots (2.5 mg) of the proteins were lyophilized and stored at −20 °C.

A total of 2.5 mg of the lyophilized protein was dissolved in 1.2 ml of 4 M guanidinium-HCl, 1.7 mM HCl, pH 1.6. Spin-label conjugation was carried out by dissolving 0.5 mg of 3-iodomethyl-(1-oxy-2,2,5,5-tetramethylpyrrolidine) (catalogue number I709500; Toronto Research Chemicals) in 10 μl of ethanol, followed by the addition of 140 μl of 1 M Tris-HCl, pH 8, and adding the resulting mixture to the protein solution. After incubation for 1 h at room temperature, 30 μl of 1 M dithiothreitol was added and the incubation continued for another 1.5 h. The sample was loaded onto a Superdex-75 column (1.6 × 60 cm, GE Healthcare) equilibrated in 4 M guanidinium-HCl, 20 mM sodium formate, pH 2.6, at a flow rate of 1.4 ml min<sup>−1</sup> at room temperature. Peak fractions were pooled and the concentration was estimated by measuring absorbance at 280 nm. The extent of labelling was 100% as determined by MALDI-TOF analysis on a Voyager-DE instrument (Perceptive Biosystems). Spin-labelling does not perturb the structure of the <sup>SFNF</sup>PR(D25N) mini-precursor as judged by NMR spectroscopy. The three mutations within the protein core, T12C, E34C and V82C, are frequently mutated in viable HIV-1 variants and are therefore not expected to alter significantly the catalytic properties of the protease<sup>2</sup>. It should be noted that V82C is located close to the substrate binding cleft comprising residues 80–81 and 83–85, but its side chain points outwards towards solvent. In the one instance where kinetic data are available for a mutation at position 82 (V82A), only a modest 10–15% decrease in *k*<sub>cat</sub>/*K*<sub>m</sub> relative to wild type is observed, and structural differences between wild-type protease and the V82A mutant are insignificant, with an r.m.s. deviation between the two crystal structures of only 0.12 Å for all main chain atoms<sup>27</sup>. Thus, the presence of a bulky spin label at position 82 would not be expected to result in a major perturbation in catalytic activity.

After extensive dialysis against 7 mM HCl, 1.4 mg each of the conjugated protein and the U-[<sup>2</sup>H/<sup>13</sup>C/<sup>15</sup>N]-labelled <sup>SFNF</sup>PR(D25N) protein were mixed and adjusted to a final concentration of 0.25 mg ml<sup>−1</sup> protein, 35% acetonitrile and 0.05% trifluoroacetic acid. The solution was dialysed against 2 l of 7 mM HCl and 4 l of 20 mM sodium phosphate, pH 5.8, each for a period of 1.5–2 h and concentrated to ~400 μM using Amicon Ultra-4 (10,000 MWCO) devices. Protein concentration (mg ml<sup>−1</sup>) was determined spectrophotometrically using  $\epsilon$  (0.1%) = 1.097 at 280 nm.

**Control active <sup>SFNF</sup>PR(D25) mini-precursor protease construct.** The <sup>SFNF</sup>PR(D25N) precursor construct does not undergo autoprocessing owing to the substitution of the active site Asp 25 by Asn. To verify that <sup>SFNF</sup>PR(D25N) represents a suitable model system we examined the autoprocessing activity of the corresponding <sup>SFNF</sup>PR(D25) precursor; that is, the precursor without the active site mutation. Most of the expressed protein undergoes maturation at the N terminus (between Phe-Pro) of the protease in the control <sup>SFNF</sup>PR(D25) precursor to produce the mature protease as expected. This was confirmed by

subjecting an aliquot of the purified (dissolved) inclusion bodies to electrospray-mass spectrometry. The measured mass of 10,728 Da clearly corresponds to the PR(D25) mature protease (calculated mass of 10,728.3 Da). Under identical conditions of analysis for <sup>SFNF</sup>PR(D25N), which is devoid of catalytic activity, only the full-length protein corresponding to a mass of 11,222 Da (calculated mass of 11,222.8 Da) is observed consistent with previous observations from studies using the inactive full-length TFR–PR(D25N) precursor, which does not undergo maturation<sup>5</sup>, as compared to the active TFR–PR(D25) precursor, which exhibits time-dependent processing at the p6<sup>pol</sup>–protease junction to release the mature protease<sup>9,10</sup>.

**NMR experiments.** All NMR data were acquired at 20 °C on a Bruker DRX600 spectrometer equipped with a z-gradient triple resonance cryoprobe.

Measurement of translational diffusion coefficients (*D*<sub>s</sub>) by pulse field gradient NMR<sup>28</sup> was carried out using the Watergate BPP-LED pulse scheme described previously<sup>29</sup>. The translational diffusion coefficient *D*<sub>s</sub> is derived from a linear least-squares fit to a plot of  $\ln[I(f)/I(f_0)]$  versus (*f*<sup>2</sup> − *f*<sub>0</sub><sup>2</sup>):

$$\ln[I(f)/I(f_0)] = -(\gamma\delta G_{\max})^2(f^2 - f_0^2)(A - \delta/3 - \tau/2)D_s$$

where *I*(*f*) and *I*(*f*<sub>0</sub>) are the intensities of the NMR signal at fractional gradient strengths of *f* and *f*<sub>0</sub>; *f*<sub>0</sub> is the fractional gradient strength of the reference spectrum (0.1); *f* is the fractional gradient strength with values of 0.2, 0.3, 0.4, 0.5 and 0.6 times *G*<sub>max</sub>, the maximum gradient strength (70 × 10<sup>−4</sup> T cm<sup>−1</sup>);  $\gamma$  is the gyromagnetic ratio of <sup>1</sup>H (2.6752 × 10<sup>8</sup> s<sup>−1</sup> T<sup>−1</sup>); *A* = 15.4 ms;  $\delta$  = 5 ms (gradient duration); and  $\tau$  = 0.2 ms. The overall diffusion delay is 10 ms. The value of the scaling factor  $(\gamma\delta G_{\max})^2(A - \delta/3 - \tau/2)$  is 1.19 × 10<sup>10</sup> s m<sup>−2</sup>. The values of *D*<sub>s</sub> were 9.3(±0.4) × 10<sup>−11</sup> and 12.9(±0.5) × 10<sup>−11</sup> m<sup>2</sup> s<sup>−1</sup> for the <sup>SFNF</sup>PR(D25N) precursor and the mature PR(D25N) dimer, respectively, at the same (0.4 mM) subunit concentration (Supplementary Fig. 1d). The ratio of the two *D*<sub>s</sub> values (0.72 ± 0.04) is fully consistent with the expected value of 0.75 for a *D*<sub>s</sub><sup>monomer</sup>/*D*<sub>s</sub><sup>dimer</sup> ratio<sup>28</sup>, placing an upper limit of about 10% for the population of dimeric species.

<sup>15</sup>N-<sup>1</sup>H heteronuclear NOE measurements were carried out using a flip-back scheme as described<sup>30</sup>. Residues −4 to 9 and 95–99 of <sup>SFNF</sup>PR(D25N) have heteronuclear <sup>15</sup>N-<sup>1</sup>H NOE values ranging from −1 to 0.5 indicating that they are disordered and highly mobile. Backbone assignments were derived using the following three-dimensional triple resonance experiments: HNCO, HN(CO)CA and CBCA(CO)NH<sup>31,32</sup>. The weighted mean backbone chemical shift difference between different constructs is given by  $[\Delta\delta_{\text{HN}}^2 + \Delta\delta_{\text{N}/25}^2 + \Delta\delta_{\text{C}\alpha/4}^2]^{1/2}$  as described previously<sup>33</sup>. A comparison of <sup>1</sup>H/<sup>15</sup>N/<sup>13</sup>Cα chemical shifts reveals significant perturbations relative to the corresponding mature dimeric PR(D25N) for residues located at the dimer interface (Supplementary Fig. 1b), but only minor perturbations relative to the equivalent monomeric PR(1–95) construct obtained by deletion of the C-terminal four residues (Supplementary Fig. 1c)<sup>5</sup>. Analysis of the chemical shift index (based on <sup>13</sup>Cα, <sup>13</sup>Cβ and <sup>13</sup>C′ shifts)<sup>34</sup> for <sup>SFNF</sup>PR(D25N) and PR(D25N) indicates that the secondary structure elements are preserved in the precursor with the exception of the N- and C-terminal strands which form an intersubunit four-stranded antiparallel β-sheet in the mature dimer (Supplementary Fig. 1e).

PRE <sup>1</sup>H<sub>N</sub>–<sup>15</sup>N rates are given by the difference in *R*<sub>2</sub> relaxation rates between the paramagnetic (spin-labelled) and diamagnetic states of the protein. *R*<sub>2</sub> rates were determined from a two-time-point interleaved two-dimensional <sup>1</sup>H–<sup>15</sup>N correlation-based experiment, as described previously<sup>21</sup>. The time interval between the two time points was 32 ms for the intermolecular PRE measurements and 4 ms for the intramolecular PRE measurements. The short time interval for the latter is used to minimize any errors in *R*<sub>2</sub> rates introduced by any potential diamagnetic contamination (that is, spin-labelling less than 100%)<sup>21</sup>.

**Tertiary structure of <sup>SFNF</sup>PR(D25N).** To verify that the tertiary structure of the ordered region of <sup>SFNF</sup>PR(D25N) (that is, residues 10–94) is the same as that of an individual subunit of the mature protease, we made use of the CS-Rosetta chemical shift structure determination algorithm which uses a hybrid approach of chemical-shift-based fragment selection and ROSETTA Monte Carlo driven fragment assembly<sup>35</sup>. The resulting ten lowest energy models are essentially identical to the corresponding region of the mature dimer with a backbone r.m.s. deviation of only 1.3 ± 0.2 Å (Supplementary Fig. 1e).

**PRE calculations and ensemble refinement.** Because the electron relaxation rate  $\tau_e$  of the free radical is much longer than that of the protein rotational correlation time  $\tau_r$ <sup>21</sup>, the PRE correlation time  $\tau_c$  [ $= (\tau_r^{-1} + \tau_e^{-1})^{-1}$ ] for the calculation of intermolecular PRE rates was assumed to be the same as  $\tau_r$  (12 ns) for the mature protease dimer<sup>36</sup>. To account for the flexibility of the linker between the spin label and the protein backbone, a ten-conformer randomized ensemble was used to represent the conformational space sampled by the spin label. The randomized ensemble was generated by high-temperature simulated annealing and slow cooling in Xplor-NIH<sup>22</sup> subject to a target function compris-

ing stereochemical terms, a quartic van der Waals repulsion term to prevent atomic overlap between the spin label and the protein, and a multidimensional conformational database potential of mean force<sup>37</sup> describing the  $\phi/\psi/\chi_1$  conformational space available to the surface cysteine residue to which the spin label was conjugated. Note that overlap between the members of the Cys spin-label ensemble is permitted as the ten-member ensemble represents a distribution of states. To ensure full sampling of the conformational space available to the spin label a different ten-conformer randomized ensemble was used for each structure calculation. Agreement between observed and calculated  $\Gamma_2$  rates is given by the PRE Q-factor,  $Q_{\text{PRE}}$ :<sup>23</sup>

$$Q_{\text{PRE}} = \left[ \sum_i \left\{ \Gamma_{2,i}^{\text{obs}} - p \langle \Gamma_{2,i}^{\text{calc}} \rangle \right\}^2 / \sum_i \left( \Gamma_{2,i}^{\text{obs}} \right)^2 \right]^{1/2}$$

where  $\Gamma_{2,i}^{\text{obs}}$  and  $\langle \Gamma_{2,i}^{\text{calc}} \rangle$  are the observed and ensemble average calculated transverse  $\Gamma_2$  rates for residue  $i$ , respectively, and  $p$  is the overall population of the encounter complex species. All members of an ensemble of size  $N_e$  are weighted equally. For the average Q-factor  $\langle Q \rangle$  for all calculated  $n$  ensembles,  $\langle \Gamma_{2,i}^{\text{calc}} \rangle$  is averaged over the members of each  $N_e$  ensemble. For the ensemble of ensembles average PRE Q-factor,  $Q_{\text{ee}}$ ,  $\langle \Gamma_{2,i}^{\text{calc}} \rangle$  is averaged over all ensemble members and all ensembles<sup>14</sup>.

The coordinates used in the Xplor-NIH<sup>22</sup> calculations were taken from the X-ray structure of the unliganded mature HIV-1 protease dimer (Protein Data Bank accession code 1HHP)<sup>24</sup>. Residues 10–94 were treated as a rigid body, and the flexible N- and C-terminal residues were not included in the calculations. The coordinates of the isotopically labelled subunit were held fixed, the initial positions of the spin-labelled subunit (at natural isotopic abundance) were randomized, and rigid-body simulated annealing was carried out against the PRE data sets for the spin label conjugated to the T12C, E34C and V82C sites simultaneously. The target function comprises a PRE restraint term<sup>23</sup>, a quartic van der Waals repulsion term to prevent atomic overlap between the spin-labelled and isotopically labelled subunits, and a very weak radius of gyration term<sup>38</sup> to ensure that each member of the ensemble makes at least some intermolecular contacts<sup>14,39</sup>. Note that atomic overlap between ensemble members of spin-labelled subunits is permitted as these represent separate but rapidly interconverting configurations of the encounter complex species<sup>14,39</sup>. A grid search was performed varying the population of heterodimer and the ensemble size  $N_e$  used to represent the self-associated species<sup>14</sup>. For each ensemble size and population of encounter complex species, 100 calculations were carried out. Ensembles were ranked by PRE Q-factor and van der Waals repulsion energies, and the top 20 ensembles with the smallest PRE Q-factors were used for subsequent analysis<sup>39</sup>. Structures were rendered using PyMol (<http://www.pymol.org>) and re-weighted atomic probability density maps were generated using Xplor-NIH<sup>22</sup> as described<sup>25</sup>.

**d.r.m.s. metric.** One metric we used to compare the precursor encounter complexes with the mature dimer was the distance root mean square (d.r.m.s.) metric defined by<sup>40</sup>:

$$\text{d.r.m.s.} = \frac{1}{N} \sum_{i,j} \left| d_{i,j}^{\text{precursor}} - d_{i,j}^{\text{mature}} \right|$$

where  $N$  is the number of distinct residue pairs ( $i, j$ ), and  $d_{i,j}^{\text{precursor}}$  and  $d_{i,j}^{\text{mature}}$  are

the distance matrices in a calculated precursor encounter complex structure and the mature HIV-1 protease dimer structure, respectively.

**Spherical coordinate systems used to describe relative subunit orientation in the encounter complexes.** Two spherical coordinate systems are used to describe the relative orientation of the subunits in the precursor encounter complexes<sup>40</sup>. The first (polar angle  $\phi$  and azimuth angle  $\theta$ ) describes the orientation of the vector joining the centre of masses of the two subunits (shown as grey spheres in Supplementary Fig. 3a) to an external axis system with the  $z$  axis corresponding to the  $C_2$  symmetry axis of the mature dimer. The second (polar angle  $\alpha$  and azimuth angle  $\beta$ ) describes the orientation of a vector joining the centre of mass of the second subunit to an arbitrarily chosen atom of the same subunit (C $\alpha$  atom of Gly 51) relative to an axis system with the  $z'$  axis given by the vector joining the centre of masses of the two subunits (with the red subunit in Fig. 2 corresponding to the fixed reference subunit) (Supplementary Fig. 3a).

26. Wondrak, E. M. & Louis, J. M. Influence of flanking sequences on the dimer stability of human immunodeficiency virus type 1 protease. *Biochemistry* **35**, 12957–12962 (1996).
27. Mahalingam, B. *et al.* Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site. *Eur. J. Biochem.* **271**, 1516–1524 (2004).
28. Altieri, A. S., Hinton, D. P. & Byrd, R. A. Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. *J. Am. Chem. Soc.* **117**, 7566–7567 (1995).
29. Chou, J. J., Baber, J. L. & Bax, A. Characterization of phospholipid mixed micelles by translational diffusion. *J. Biomol. NMR* **29**, 299–308 (2004).
30. Grzesiek, S. & Bax, A. The importance of not saturating  $H_2O$  in protein NMR: application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593–12594 (1993).
31. Clore, G. M. & Gronenborn, A. M. Two-, three- and four-dimensional NMR methods for obtaining larger and more precise three-dimensional structures of proteins in solution. *Annu. Rev. Biophys. Chem.* **20**, 29–63 (1991).
32. Clore, G. M. & Gronenborn, A. M. Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Methods Enzymol.* **239**, 349–363 (1994).
33. Grzesiek, S., Stahl, S. J., Wingfield, P. T. & Bax, A. The CD4 determinant of downregulation by HIV-1 Nef directly binds to Nef: mapping of the Nef binding surface by NMR. *Biochemistry* **35**, 10256–10261 (1996).
34. Wishart, D. S. & Sykes, B. D. The  $^{13}C$  chemical-shift index: a simple method for the identification of protein secondary structure using  $^{13}C$  chemical-shift data. *J. Biomol. NMR* **4**, 171–180 (1994).
35. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl Acad. Sci. USA* **105**, 4685–4690 (2008).
36. Katoh, E. *et al.* A solution NMR study of the binding kinetics and internal dynamics of an HIV-1 protease-substrate complex. *Protein Sci.* **12**, 1376–1385 (2003).
37. Clore, G. M. & Kuszewski, J.  $\chi_1$  rotamer populations and angles of mobile surface side chains are accurately predicted by a torsion angle database potential of mean force. *J. Am. Chem. Soc.* **124**, 2866–2867 (2002).
38. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J. Am. Chem. Soc.* **121**, 2337–2338 (1999).
39. Tang, C., Ghirlando, R. & Clore, G. M. Visualization of transient ultra-weak protein self-association in solution using paramagnetic relaxation enhancement. *J. Am. Chem. Soc.* **130**, 4048–4056 (2008).
40. Kim, Y. C., Tang, C., Clore, G. M. & Hummer, G. Replica exchange simulations of transient encounter complexes in protein-protein association. *Proc. Natl Acad. Sci. USA* **105**, 12855–12860 (2008).



# Biochemistry's new look

**Until now, metabolomics researchers have had to adapt technology developed mainly for proteomics. But there are now solutions designed with them in mind. Nathan Blow reports.**

Metabolomics — the comprehensive study of metabolic reactions — is gaining ground alongside its older siblings genomics and proteomics. “Unlike some of the other ‘omics’ that we have seen, metabolomics is going to produce a lot of useful information right from the start,” says Gary Siuzdak, professor of molecular biology at the Scripps Research Institute in La Jolla, California. He is one of a growing number of biologists using advanced technology to explore biochemical questions on a scale that would have seemed impossible a decade ago.

“The metabolome is the best indicator of an organism’s phenotype,” says David Wishart at the University of Alberta in Edmonton, Canada. Wishart was one of the instigators of the Human Metabolome Project, a US\$7.5-million effort funded by Genome Canada to systematically characterize the metabolites of the human body. He gives the example of a person holding their breath for five minutes. Although genomic or proteomic analysis would not provide any evidence of stress during this short period — even as the person turns blue — metabolite profiles would show dramatic changes within the body.

Unlike a genome or even a proteome, however, a metabolome is tricky to pin down. Wishart notes that although researchers know there are 3 billion base pairs in the human genome, if you ask biochemists how many small-molecule metabolites there are in the human body, they come back with numbers ranging from 3,000 to 100,000. And this poses a real challenge for metabolomics research, as both ends of the scale could be correct.

The Human Metabolome Project has pegged the number of endogenous metabolites in the human body at around 3,000 — which most researchers agree on. But humans also take in small molecules from the environment — preservatives in food, chemicals in the air, metabolites produced through the breakdown of drugs and toxins — making an exact figure hard to determine.

## Separation anxiety

With metabolites in such a state of flux, researchers do not have an easy task. Nevertheless, advances in chromatography, mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) are allowing them to make headway in defining different



**The Pegasus 4D GCxGC MS TOF system enables multidimensional approaches to GC separation.**

metabolomes and understanding how changes in the concentrations of metabolites relate to human health and disease.

One of the problems is that metabolites come in a variety of chemical forms. “I would say one of the real challenges of metabolomics is that each metabolite is its own unique puzzle,” says Trent Northen, a scientist at Lawrence Berkeley National Laboratory in California. And in most cases the first step in solving the puzzle is isolating the metabolite for analysis.

No one separation method works for all metabolites, so researchers rely on combinations of gas chromatography (GC), liquid chromatography (LC) and emerging capillary electrophoresis (CE).

Historically, GC separation has had the edge. “GC–MS technology may not be sexy,

but huge databases are available,” says Northen. These GC–MS databases, compiled over more than four decades, enable researchers to compare a wide range of spectra to arrive at a chemical identification.

Multidimensional GC, often called ‘GCxGC’ or two-dimensional GC, offers even better separation. “When people are doing GCxGC, they are trying to get more separation chroma-

tographically of very complex samples,” says Steven Fischer, a senior applications chemist at Agilent Technologies in Santa Clara, California. To achieve this, GCxGC uses two separation phases, such as a non-polar and a polar phase, in two capillary columns in series in the instrument.

Agilent recently introduced the 7890 GC system, which can perform multidimensional GC, and Thermo Fisher Scientific in Waltham Massachusetts, has developed the Trace GCxGC system. The Pegasus 4D GCxGC MS time-of-flight (TOF) system from LECO, based in St Joseph, Michigan, uses a thermal modulator placed between the two GC columns to collect effluent from the first column before going into the second phase of separation. The power of the multidimensional approach is starting to be reported. In May this year a group reported the use of GCxGC with LC–MS to generate a draft metabolic network for the single-celled alga *Chlamydomonas reinhardtii*<sup>1</sup>.

## Class action

GC is particularly useful for mixtures of volatiles, such as steroids, saccharides and sugar alcohols, which can be sent directly into the gas phase for separation. Metabolites in human biofluids and tissues therefore present a technical challenge for GC, as most are not volatile. Non-volatile metabolites either need complicated chemical transformation before GC or



**Gary Siuzdak: developing new approaches to metabolite identification.**

separation by other types of chromatography.

One of these is high-performance liquid chromatography (HPLC), a well-established lab workhorse. It uses a combination of solvents, pressure and matrix particle sizes to separate molecules on the basis of their retention times in a column packed with matrix. HPLC can separate a broad range of metabolites, including non-volatiles, and remains a favourite among metabolomics researchers.

Most advances in HPLC involve increases in the pressure applied and changes in matrix particle size. Ultra-performance liquid chromatography (UPLC), commercialized by Waters Corporation in Milford, Massachusetts, is becoming more widely used in the metabolomics community. It takes advantage of higher pressure (83 megapascals compared with 21 megapascals for HPLC) and smaller particles (less than 2 micrometres diameter compared with 3 micrometres for HPLC) to obtain faster separation times.

But like GC, HPLC has technical stumbling blocks. Reversed-phase HPLC (in which the stationary phase is non-polar) is often used for metabolomics analysis, but reversed-phase separation often fails with hydrophilic metabolites. These tend to be so water soluble that they interact poorly with the non-polar bonding phase and are

rapidly eluted, according to Phil Koerner, a senior technical manager from chromatography specialists Phenomenex in Torrance, California.

So in 2007, Phenomenex introduced the Luna HILIC column. "I like to refer to it as reverse reverse-phase chromatography," says Koerner. In the HILIC approach, the weak solvent, which is applied first, is a polar organic solvent (not water as in reversed-phase HPLC), and the strong solvent, applied second, is water. This causes the order of elution to be completely reversed, with the most hydrophilic compounds being eluted last. Although Koerner acknowledges that the HILIC approach is not new, it was the need to separate hydrophilic metabolites on the large scale required by metabolomics that led Phenomenex and other companies, such as Waters and Tosoh Bioscience of Stuttgart, Ger-

many, to start supplying a greater number and range of HILIC columns.

Capillary electrophoresis followed by MS (CE-MS) is not yet so popular with the metabolomics community as either GC or HPLC, but several developers are hoping to change this. "It can be very difficult to use this approach," acknowledges Ryuji Kanno, president of Human Metabolome Technologies based in Tokyo, Japan. This approach uses electrophoretic mobility to separate low-molecular-weight ionic compounds that are difficult to separate by GC or HPLC. The company has been working closely with Agilent to develop optimized reagents and capillary columns, and is providing training with Agilent's CE-qTOF MS system to make the CE approach more accessible to metabolomics researchers, says Kanno.

Mass spectrometry is not the only method

that can be used to detect metabolites once separated. Wishart and his colleagues recently compared MS and NMR to look at metabolites in cerebral spinal fluid<sup>2</sup>. They found little overlap in the metabolites detected by the two methods, and the conclusion was clear: "We do not have a single perfect metabolite detector," says Wishart.

MS and NMR each have their supporters. "One of the main strengths of NMR is that



HILIC columns are making the hunt for hydrophilic metabolites easier.

PHENOMENEX

## DARK MATTER

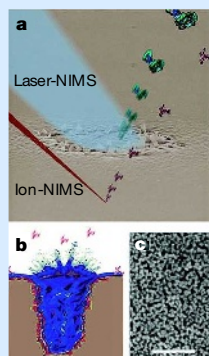
"There are a lot of small molecules that we do not even know about yet," says Arthur Castle, programme director for the Roadmap Metabolomics Technology development programme at the US National Institutes of Health in Bethesda, Maryland. Metabolomics has a good handle on analysing human primary metabolites, but when it comes to lipids, secondary metabolites, xenobiotics and the products of gut microflora, we are just scratching the surface, says Castle.

The problem is part technological, part informatics. Steven Fischer at Agilent Technologies in Santa Clara, California, points out that some compounds are not stable and undergo chemical transformation during separation. Researchers will probably still see these transformed molecules by mass spectrometry (MS), but they may

be misidentified, highlighting the need for follow-up experiments. And this is where metabolomics has an advantage over the other 'omics'. "There is so much knowledge of biochemistry that when we find a potential biomarker or a new drug mechanism we already know a lot about it," says Michael Milburn, chief scientific officer at Metabolon in Durham, North Carolina.

Trent Northen, now at the Lawrence Berkeley Laboratory in Berkeley, California, and Oscar Yanes, working in Gary Siuzdak's lab at the Scripps Research Institute in La Jolla, California, may have developed a new way to get at some of these 'unknown' molecules with

an ionization technique called nanostructure-initiator mass spectrometry (NIMS)<sup>4</sup>. The idea is to transfer a biomolecule into



NIMS: nanostructured surface releases material.

the gas phase from a nanostructured surface simply by making that surface disappear. "We came up with the idea of putting a wax underneath, so when the nanostructured surface was irradiated it would melt and vaporize, allowing the molecule to go into the gas phase," says Northen. They finally came up with a perfluorinated surface for the trapped initiator phase.

Siuzdak's group found that NIMS worked not only for proteins, but also for small molecules such as metabolites. "Perfluorinated do not ionize well, so it allows us to

see things in the lower mass region where metabolites like to hang out," says Siuzdak.

Using perfluorinated materials may have another advantage for metabolomics. "As these are highly hydrophobic surfaces, we can apply very dirty complex samples, such as blood, for direct analysis," says Yanes. If you put a drop of urine or blood on the NIMS chip the metabolites will attach, but all the salts and other chemicals that normally interfere with MS stay in solution. Yanes is now exploiting this property to follow drug metabolism by looking at uptake in blood, clearance in urine and tissue localization.

For Siuzdak, exploring the 'unknown' metabolite world is an exciting prospect. "We are getting involved in an area where we don't know what the molecules' structures are or what they do, so it is really just a fantastic area for discovery."

REF. 4

N.B.





The maxis system from Bruker Daltonics can use both UPLC and CE separation approaches.

Newton. He says researchers at Chenomx have performed many studies in which biologically meaningful differences between samples were easily captured with NMR, even though some compounds in the samples probably fell below the sensitivity limits of the instrument (see 'Dark matter').

MS, on the other hand, is a very sensitive method for metabolite identification and, unlike NMR, is easily coupled to upstream separation techniques. Siuzdak says his group can see thousands of molecules in an MS analysis — and that number can be doubled by changing from positive- to negative-ion mode. And by using both reversed-phase chromatography and HILIC columns, they are seeing more hydrophilic compounds in their analyses than before. "I would venture that we are now seeing over an order of magnitude more than what you would see with NMR," he says.

### Detector development

As researchers in the MS camp turn towards TOF and ion-trap MS instruments for metabolite analysis, developers are responding to their complex needs. Bruker Daltonics in Billerica, Massachusetts, has introduced the maxis ultra-high resolution (UHR)-TOF MS system, which can accommodate both UPLC and CE separation. Applied Biosystems in Foster City, California, in collaboration with MDS SCIEX in Toronto, Ontario, have the ion-trap system 4000 QTrap LC/MS/MS that can interface with Applied Biosystem's LightSight software for small-molecule analysis and identification.

it is an unbiased, universal detector," says Jack Newton, a product manager at Chenomx in Edmonton, Canada, which was co-founded by Wishart in 2000. This attribute, along with NMR's ability to determine structure and perform quantitative analysis is particularly attractive to metabolomics researchers who need a way to compare and exchange results between labs. "The move is afoot — people want to get

to that common language of compound names and concentrations," says Wishart, as this will make integrating data sets and obtaining systems-level views of cell physiology possible.

The challenge with NMR is instrument sensitivity — NMR is less sensitive than MS, often identifying far fewer metabolites in the same sample. "For us, the relevant question is how sensitive do you need to be," says

## WINE-OMICS

For Kirsten Skogerson at the University of California, Davis, wondering about how chemical composition affects the flavour and body of a wine took her from a degree in viticulture and enology into metabolomics research. When Skogerson arrived in Oliver Fiehn's lab as a postgrad she looked for a project that would marry Fiehn's expertise in metabolomics and her interest in wine.

"There are so many questions in wine science that you could start to answer by doing a global analysis," she says. A deeper understanding of the biochemistry of grape-juice fermentation could help the winemaking industry by complementing the arts of the traditional wine taster. So Skogerson and Fiehn set out to survey wine 'metabolomes', in search of key chemical components contributing to body.

Using proton nuclear magnetic resonance (NMR)

and gas chromatography-mass spectrometry (GC-MS), they looked at 17 different white wines with a wide range of body. For GC-MS analysis, they first removed the alcohol under reduced pressure and then ran samples on a LECO Pegasus IV GC TOF MS system and analysed the spectra using the BinBase program developed in Fiehn's lab. Each wine was also directly analysed on a Bruker Daltonics 600 MHz NMR instrument with the resulting peaks being compared to the commercially available Chenomx NMR database for metabolite identification. "When you think about it, you have the grape metabolome being acted on by the yeast, plus the added complexity from the yeast metabolome, so the metabolite profile of a wine is very complex," says Skogerson.

They found a total of 413 metabolites among the wines — probably only a small fraction of



the wine metabolome — of which 108 could be positively identified. And in both data sets, the amino acid proline showed a positive correlation with body as assessed by trained wine tasters. How proline relates to body is not yet clear, however. "That is the hard part of

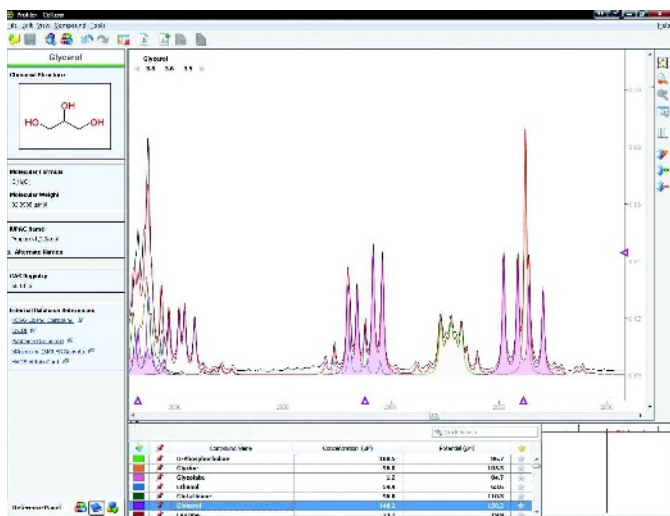
being in metabolomics — you get clues, but the follow-up is the real challenge," says Skogerson. Still, she thinks proline could be used as marker for a wine's viscosity.

Red-wine drinkers have not been forgotten. Bruker Daltonics in Billerica, Maryland, has profiled red wines for important polyphenolic secondary metabolites such as tannins, flavonoids and anthocyanins. This demonstration used the Acquity ultra-performance liquid chromatography system from Waters to separate red wine metabolites for analysis by Bruker's LC-ESI QTOF MS instrument as well as analysis by NMR coupled with Bruker's BioSpin Spectral Base analysis package.

Does knowing the chemistry behind that wonderful bottle of wine take away from the pleasure? Not according to Skogerson. "Science has the potential to bring the art of winemaking to a higher level."

N.B.





Chenomx has developed a searchable NMR database for metabolomics.

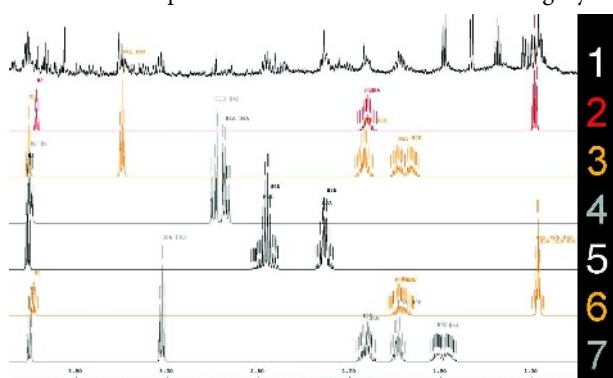
Both Agilent and Thermo Fisher Scientific also offer MS systems and software packages designed for metabolite analysis.

Some researchers and developers are designing platforms to bring the two camps closer together — incorporating NMR and MS instruments in a single system. Bruker BioSpin in Billerica, Massachusetts, has developed the Metabolic Profiler, a system that combines a liquid handler, the Avance III NMR spectrometer and an LC-electrospray ionization (ESI)-microTOF MS, all under the control of a single data-management and analysis system.

But what researchers dream of is a single detection ‘chip’ for all metabolites. “In my lab we have four platforms, and each platform looks at a certain part of the metabolome,” says Oliver Fiehn, a metabolomics researcher at the University of California, Davis. But he doubts that a single chip could ever become reality.

“The lack of such a technology is the Achilles heel of metabolomics,” says Wishart, noting that the most that researchers can analyse at any one time with current technologies is 10–15% of the entire metabolome — and even that’s stretching it.

“The big bottleneck is really compound identification,” says Fiehn. Unblocking it will need the addition of many more well-annotated reference spectra in the databases.



Comparisons of samples to reference spectra databases help reveal the identity of metabolites.

And that will take time. Chenomx was founded with the aim of developing a database for NMR analysis, and that has taken several years of intensive effort, says Newton. Different chemical environments can influence a compound’s NMR spectra, so researchers at Chenomx had to acquire spectra at ten pHs, ranging from 4 to 9, for each of the more than 300 reference compounds now in their proprietary database.

Metabolite data-

bases for MS have also been springing up as more researchers move into the field. One of the first was METLIN (<http://metlin.scripps.edu>), a publicly accessible database that was started in Siuzdak’s lab.

“We currently have 23,000 metabolites in there,” says Siuzdak, of which around 2,500 are identified endogenous metabolites. METLIN also contains a set of about 8,000 theoretical di- and tripeptides along with theoretical lipids, drugs and metabolites.

To expand the scope of METLIN, Agilent has collaborated with the Scripps Center for Mass Spectrometry to analyse chromatographic standards and add information about mass and retention time, with the intent of using these properties in addition to isotope pattern matching for identification. “Our goal is to get to the point where the most common metabolites encountered by researchers are easily identifiable,” says Agilent’s Fischer.

Gregory Stephanopoulos, a chemical engineer at the Massachusetts Institute of Technology in Cambridge, is taking a different approach to metabolite identification. Several years ago a student approached him with an interesting metabolomics project, but the catch was that the lab would first have to increase the number of reference spectra in its library to enable metabolite identification. Although Stephanopoulos liked the project, he did not like the idea of simply collecting spectra to fill a data-

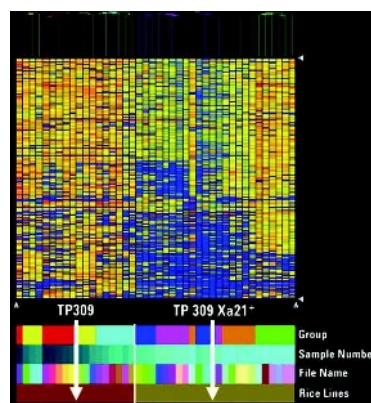
base. “I thought that there had to be a better way to deal with the issue,” he recalls.

The result is a web-based program called SpectConnect, which was launched in 2007 to help researchers identify important metabolites that might serve as biomarkers<sup>3</sup>. The SpectConnect algorithm tracks and catalogues GC–MS spectra that are conserved in multiple samples — an indication that these represent real compounds instead of noise in the sample, helping to guide researchers with their follow-up efforts at full metabolite identification.

## Numbers game

The good news for metabolomics researchers is that NMR and MS metabolite databases are increasing in both number and size as new metabolomes are analysed. “One of the things that changed for us over the past 18 months is the places we are applying the technology,” says Michael Milburn, chief scientific officer at Metabolon in Durham, North Carolina. Metabolomic approaches are now addressing biological questions in areas ranging from drug discovery and cosmetics development to plant

science and winemaking (see ‘Wine-omics’). Publicly accessible databases include MassBank for high-resolution ESI mass spectra of metabolites ([www.massbank.jp](http://www.massbank.jp)), BinBase for processing and analysing of dissimilar MS spectra (<http://sourceforge.net/projects/binbase/>), and MetWare (<http://msbi.ipb-halle.de/msbi/metware/>) for the storage and analysis of metabolomic experiments. Commercial databases include Metabolon’s,



Samples can be grouped by the similarity of mass abundance profiles.

containing spectra of more than 6,000 reference metabolites, and Bio-Rad’s KnowITAll spectral database of more than 1.3 million entries, including MS and NMR references.

But there’s still a way to go before metabolite identification is as simple as ‘query and get a chemical name’. “The database changes have been encouraging,” says Stephanopoulos. But not enough to change his mind about the need for tools such as SpectConnect.

Arthur Castle, programme director for the Roadmap Metabolomics Technology development programme at the US National Institutes of Health, has seen the pieces falling into place over the past couple of years. “The technology is very close to being there — it is just a question of putting it all together now,” he says.

**Nathan Blow is technology editor for *Nature* and *Nature Methods*.**

1. May, P. et al. *Genetics* **179**, 157–166 (2008).
2. Wishart, D. S. et al. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **15**, 164–173 (2008).
3. Styczynski, M. P. et al. *Anal. Chem.* **79**, 966–973 (2007).
4. Northen, T. R. et al. *Nature* **449**, 1033–1036 (2007).

separation by other types of chromatography.

One of these is high-performance liquid chromatography (HPLC), a well-established lab workhorse. It uses a combination of solvents, pressure and matrix particle sizes to separate molecules on the basis of their retention times in a column packed with matrix. HPLC can separate a broad range of metabolites, including non-volatiles, and remains a favourite among metabolomics researchers.

Most advances in HPLC involve increases in the pressure applied and changes in matrix particle size. Ultra-performance liquid chromatography (UPLC), commercialized by Waters Corporation in Milford, Massachusetts, is becoming more widely used in the metabolomics community. It takes advantage of higher pressure (83 megapascals compared with 21 megapascals for HPLC) and smaller particles (less than 2 micrometres diameter compared with 3 micrometres for HPLC) to obtain faster separation times.

But like GC, HPLC has technical stumbling blocks. Reversed-phase HPLC (in which the stationary phase is non-polar) is often used for metabolomics analysis, but reversed-phase separation often fails with hydrophilic metabolites. These tend to be so water soluble that they interact poorly with the non-polar bonding phase and are

rapidly eluted, according to Phil Koerner, a senior technical manager from chromatography specialists Phenomenex in Torrance, California.

So in 2007, Phenomenex introduced the Luna HILIC column. "I like to refer to it as reverse reverse-phase chromatography," says Koerner. In the HILIC approach, the weak solvent, which is applied first, is a polar organic solvent (not water as in reversed-phase HPLC), and the strong solvent, applied second, is water. This causes the order of elution to be completely reversed, with the most hydrophilic compounds being eluted last. Although Koerner acknowledges that the HILIC approach is not new, it was the need to separate hydrophilic metabolites on the large scale required by metabolomics that led Phenomenex and other companies, such as Waters and Tosoh Bioscience of Stuttgart, Ger-

many, to start supplying a greater number and range of HILIC columns.

Capillary electrophoresis followed by MS (CE-MS) is not yet so popular with the metabolomics community as either GC or HPLC, but several developers are hoping to change this. "It can be very difficult to use this approach," acknowledges Ryuji Kanno, president of Human Metabolome Technologies based in Tokyo, Japan. This approach uses electrophoretic mobility to separate low-molecular-weight ionic compounds that are difficult to separate by GC or HPLC. The company has been working closely with Agilent to develop optimized reagents and capillary columns, and is providing training with Agilent's CE-qTOF MS system to make the CE approach more accessible to metabolomics researchers, says Kanno.

Mass spectrometry is not the only method

that can be used to detect metabolites once separated. Wishart and his colleagues recently compared MS and NMR to look at metabolites in cerebral spinal fluid<sup>2</sup>. They found little overlap in the metabolites detected by the two methods, and the conclusion was clear: "We do not have a single perfect metabolite detector," says Wishart.

MS and NMR each have their supporters. "One of the main strengths of NMR is that



HILIC columns are making the hunt for hydrophilic metabolites easier.

PHENOMENEX

## DARK MATTER

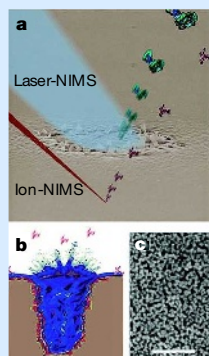
"There are a lot of small molecules that we do not even know about yet," says Arthur Castle, programme director for the Roadmap Metabolomics Technology development programme at the US National Institutes of Health in Bethesda, Maryland. Metabolomics has a good handle on analysing human primary metabolites, but when it comes to lipids, secondary metabolites, xenobiotics and the products of gut microflora, we are just scratching the surface, says Castle.

The problem is part technological, part informatics. Steven Fischer at Agilent Technologies in Santa Clara, California, points out that some compounds are not stable and undergo chemical transformation during separation. Researchers will probably still see these transformed molecules by mass spectrometry (MS), but they may

be misidentified, highlighting the need for follow-up experiments. And this is where metabolomics has an advantage over the other 'omics'. "There is so much knowledge of biochemistry that when we find a potential biomarker or a new drug mechanism we already know a lot about it," says Michael Milburn, chief scientific officer at Metabolon in Durham, North Carolina.

Trent Northen, now at the Lawrence Berkeley Laboratory in Berkeley, California, and Oscar Yanes, working in Gary Siuzdak's lab at the Scripps Research Institute in La Jolla, California, may have developed a new way to get at some of these 'unknown' molecules with

an ionization technique called nanostructure-initiator mass spectrometry (NIMS)<sup>4</sup>. The idea is to transfer a biomolecule into



NIMS: nanostructured surface releases material.

the gas phase from a nanostructured surface simply by making that surface disappear. "We came up with the idea of putting a wax underneath, so when the nanostructured surface was irradiated it would melt and vaporize, allowing the molecule to go into the gas phase," says Northen. They finally came up with a perfluorinated surface for the trapped initiator phase.

Siuzdak's group found that NIMS worked not only for proteins, but also for small molecules such as metabolites. "Perfluorinated do not ionize well, so it allows us to

see things in the lower mass region where metabolites like to hang out," says Siuzdak.

Using perfluorinated materials may have another advantage for metabolomics. "As these are highly hydrophobic surfaces, we can apply very dirty complex samples, such as blood, for direct analysis," says Yanes. If you put a drop of urine or blood on the NIMS chip the metabolites will attach, but all the salts and other chemicals that normally interfere with MS stay in solution. Yanes is now exploiting this property to follow drug metabolism by looking at uptake in blood, clearance in urine and tissue localization.

For Siuzdak, exploring the 'unknown' metabolite world is an exciting prospect. "We are getting involved in an area where we don't know what the molecules' structures are or what they do, so it is really just a fantastic area for discovery."

REF. 4

N.B.





The maxis system from Bruker Daltonics can use both UPLC and CE separation approaches.

Newton. He says researchers at Chenomx have performed many studies in which biologically meaningful differences between samples were easily captured with NMR, even though some compounds in the samples probably fell below the sensitivity limits of the instrument (see 'Dark matter').

MS, on the other hand, is a very sensitive method for metabolite identification and, unlike NMR, is easily coupled to upstream separation techniques. Siuzdak says his group can see thousands of molecules in an MS analysis — and that number can be doubled by changing from positive- to negative-ion mode. And by using both reversed-phase chromatography and HILIC columns, they are seeing more hydrophilic compounds in their analyses than before. "I would venture that we are now seeing over an order of magnitude more than what you would see with NMR," he says.

### Detector development

As researchers in the MS camp turn towards TOF and ion-trap MS instruments for metabolite analysis, developers are responding to their complex needs. Bruker Daltonics in Billerica, Massachusetts, has introduced the maxis ultra-high resolution (UHR)-TOF MS system, which can accommodate both UPLC and CE separation. Applied Biosystems in Foster City, California, in collaboration with MDS SCIEX in Toronto, Ontario, have the ion-trap system 4000 QTrap LC/MS/MS that can interface with Applied Biosystem's LightSight software for small-molecule analysis and identification.

it is an unbiased, universal detector," says Jack Newton, a product manager at Chenomx in Edmonton, Canada, which was co-founded by Wishart in 2000. This attribute, along with NMR's ability to determine structure and perform quantitative analysis is particularly attractive to metabolomics researchers who need a way to compare and exchange results between labs. "The move is afoot — people want to get

to that common language of compound names and concentrations," says Wishart, as this will make integrating data sets and obtaining systems-level views of cell physiology possible.

The challenge with NMR is instrument sensitivity — NMR is less sensitive than MS, often identifying far fewer metabolites in the same sample. "For us, the relevant question is how sensitive do you need to be," says

## WINE-OMICS

For Kirsten Skogerson at the University of California, Davis, wondering about how chemical composition affects the flavour and body of a wine took her from a degree in viticulture and enology into metabolomics research. When Skogerson arrived in Oliver Fiehn's lab as a postgrad she looked for a project that would marry Fiehn's expertise in metabolomics and her interest in wine.

"There are so many questions in wine science that you could start to answer by doing a global analysis," she says. A deeper understanding of the biochemistry of grape-juice fermentation could help the winemaking industry by complementing the arts of the traditional wine taster. So Skogerson and Fiehn set out to survey wine 'metabolomes', in search of key chemical components contributing to body.

Using proton nuclear magnetic resonance (NMR)

and gas chromatography-mass spectrometry (GC-MS), they looked at 17 different white wines with a wide range of body. For GC-MS analysis, they first removed the alcohol under reduced pressure and then ran samples on a LECO Pegasus IV GC TOF MS system and analysed the spectra using the BinBase program developed in Fiehn's lab. Each wine was also directly analysed on a Bruker Daltonics 600 MHz NMR instrument with the resulting peaks being compared to the commercially available Chenomx NMR database for metabolite identification. "When you think about it, you have the grape metabolome being acted on by the yeast, plus the added complexity from the yeast metabolome, so the metabolite profile of a wine is very complex," says Skogerson.

They found a total of 413 metabolites among the wines — probably only a small fraction of



the wine metabolome — of which 108 could be positively identified. And in both data sets, the amino acid proline showed a positive correlation with body as assessed by trained wine tasters. How proline relates to body is not yet clear, however. "That is the hard part of

being in metabolomics — you get clues, but the follow-up is the real challenge," says Skogerson. Still, she thinks proline could be used as marker for a wine's viscosity.

Red-wine drinkers have not been forgotten. Bruker Daltonics in Billerica, Maryland, has profiled red wines for important polyphenolic secondary metabolites such as tannins, flavonoids and anthocyanins. This demonstration used the Acquity ultra-performance liquid chromatography system from Waters to separate red wine metabolites for analysis by Bruker's LC-ESI QTOF MS instrument as well as analysis by NMR coupled with Bruker's BioSpin Spectral Base analysis package.

Does knowing the chemistry behind that wonderful bottle of wine take away from the pleasure? Not according to Skogerson. "Science has the potential to bring the art of winemaking to a higher level."

N.B.



COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
<b>Companies specializing in metabolomics applications</b>			
<a href="#">BIOCRATES Life Sciences</a>	Services for the identification and quantification of metabolomic biomarkers	Innsbruck, Austria	<a href="http://www.biocrates.com">www.biocrates.com</a>
<a href="#">Chenomx</a>	NMR database development for metabolite analysis	Alberta, Canada	<a href="http://www.chenomx.com">www.chenomx.com</a>
<a href="#">Human Metabolome Technologies</a>	Developing CE separation approaches for metabolite analysis	Tokyo, Japan	<a href="http://www.humanmetabolome.com">www.humanmetabolome.com</a>
<a href="#">Ingenuity Systems</a>	Software for metabolomics and pathway analysis; Ingenuity Pathway Analysis; analysis services	Redwood City, California	<a href="http://www.ingenuity.com">www.ingenuity.com</a>
<a href="#">Metanomics Health</a>	Services for metabolite profiling	Berlin, Germany	<a href="http://www.metanomics-health.de">www.metanomics-health.de</a>
<a href="#">Metabolon</a>	Contract services provider in metabolomics	Durham, North Carolina	<a href="http://www.metabolon.com">www.metabolon.com</a>
<a href="#">Molecular Biometrics</a>	Using metabolomics approaches to develop accurate diagnostic tools	Chester, New Jersey	<a href="http://www.molecularbiometrics.com">www.molecularbiometrics.com</a>
<a href="#">Umetrics</a>	SIMCA-P+ software for the analysis of metabolomic data and biomarker discovery	Kinnelon, New Jersey	<a href="http://www.umetrics.com">www.umetrics.com</a>
<a href="#">Unilever</a>	Metabolomics research in food and flavour preservation	Vlaardingen, Netherlands	<a href="http://www.unilever.com">www.unilever.com</a>
<b>NMR instrumentation</b>			
<a href="#">Cryomagnetics</a>	NMR magnets and cryostat accessories	Oak Ridge, Tennessee	<a href="http://www.cryomagnetics.com">www.cryomagnetics.com</a>
<a href="#">JEOL</a>	NMR instrumentation, mass spectrometers and accessories	Tokyo, Japan	<a href="http://www.jeol.com">www.jeol.com</a>
<a href="#">LipoScience</a>	Using NMR to explore metabolism; NMR LipoProfile test	Raleigh, North Carolina	<a href="http://www.liposcience.com">www.liposcience.com</a>
<a href="#">Oxford Instruments</a>	Dilution refrigerators, superconducting magnets, optical and spectroscopy cryostats	Oxfordshire, UK	<a href="http://www.oxinst.com">www.oxinst.com</a>
<a href="#">Process Control Technology</a>	NMR instrumentation and accessories	Fort Collins, Colorado	<a href="http://www.pctnmr.com">www.pctnmr.com</a>
<a href="#">Process NMR Associates</a>	NMR services provider; laboratory method development for NMR	Danbury, Connecticut	<a href="http://www.process-nmr.com">www.process-nmr.com</a>
<a href="#">Wilmad-LabGlass</a>	NMR sample tubes and accessories; lab glassware	Buena, New Jersey	<a href="http://www.wilmad-labglass.com">www.wilmad-labglass.com</a>
<b>Chromatography</b>			
<a href="#">Advanced Separation Technologies</a>	Analytical and preparative chromatography products; HPLC and chiral chromatography	Bellefonte, Pennsylvania	<a href="http://www.astecusa.com">www.astecusa.com</a>
<a href="#">Alcott Chromatography</a>	HPLC equipment and software, ion chromatography, hydrochromatic chromatographs	Norcross, Georgia	<a href="http://www.alcottchromatography.com">www.alcottchromatography.com</a>
<a href="#">Biotage</a>	Chromatography columns and accessories; FLASH chromatography systems	Uppsala, Sweden	<a href="http://www.biotage.com">www.biotage.com</a>
<a href="#">Capital HPLC</a>	Suppliers of HPLC and CE columns and accessories	West Lothian, UK	<a href="http://www.capital-hplc.co.uk">www.capital-hplc.co.uk</a>
<a href="#">Cecil Instruments</a>	Modular HPLC instruments, ion chromatography, UV/visible spectrophotometers	Cambridge, UK	<a href="http://www.cecilinstruments.com">www.cecilinstruments.com</a>
<a href="#">DataApex</a>	Chromatography software	Prague, Czech Republic	<a href="http://www.dataapex.com">www.dataapex.com</a>
<a href="#">GE Healthcare</a>	AKTExpress automated liquid chromatography platform	Little Chalfont, UK	<a href="http://www.gehealthcare.com">www.gehealthcare.com</a>
<a href="#">Gilson</a>	Solid-phase extraction systems, liquid chromatography	Middleton, Wisconsin	<a href="http://www.gilson.com">www.gilson.com</a>
<a href="#">Hamilton</a>	Automated solid-phase extraction systems	Reno, Nevada	<a href="http://www.hamiltoncompany.com">www.hamiltoncompany.com</a>
<a href="#">Phenomenex</a>	Supplier of chromatography columns and accessories	Torrence, California	<a href="http://www.phenomenex.com">www.phenomenex.com</a>
<a href="#">Tosoh Bioscience</a>	Liquid chromatography columns, products and accessories	Stuttgart, Germany	<a href="http://www.tosohbioscience.com">www.tosohbioscience.com</a>
<b>Mass spectrometers</b>			
<a href="#">Agilent Technologies</a>	Mass spectrometry instruments, sample preparation and software	Santa Clara, California	<a href="http://www.agilent.com">www.agilent.com</a>
<a href="#">Applied Biosystems</a>	Mass spectrometry instruments, reagents and software; ion sources	Foster City, California	<a href="http://www.appliedbiosystems.com">www.appliedbiosystems.com</a>
<a href="#">Bruker</a>	Mass spectrometers	Billerica, Massachusetts and Bremen, Germany	<a href="http://www.bdal.de">www.bdal.de</a>
<a href="#">Hitachi High-Technologies</a>	LC systems; LC/MS systems	Tokyo, Japan	<a href="http://www.hitachi-hitec.com/global">www.hitachi-hitec.com/global</a>
<a href="#">LECO</a>	Mass spectrometry instrumentation and analysis tools	St Joseph, Michigan	<a href="http://www.leco.com">www.leco.com</a>
<a href="#">MDS SCIEX</a>	Mass spectrometry systems and technology; HPLC systems	Sunnyvale, California	<a href="http://www.mdssciex.com">www.mdssciex.com</a>
<a href="#">Shimadzu</a>	Laboratory instruments including mass spectrometers and data-management systems	Kyoto, Japan	<a href="http://www.shimadzu.com">www.shimadzu.com</a>
<a href="#">Thermo Fisher Scientific</a>	Analytical instruments including mass spectrometers, lab equipment, software, services, consumables and reagents	Waltham, Massachusetts	<a href="http://www.thermo.com">www.thermo.com</a>
<a href="#">Varian</a>	Analytical instruments including nuclear magnetic resonance and magnetic resonance imaging systems, mass spectrometers, Fourier transform-infrared and X-ray crystallography	Palo Alto, California	<a href="http://www.varianinc.com">www.varianinc.com</a>
<a href="#">Waters</a>	Liquid chromatography systems; chromatography columns and chemistry products; mass spectrometry systems; laboratory informatics solutions	Milford, Massachusetts	<a href="http://www.waters.com">www.waters.com</a>

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
<b>General</b>			
<a href="#">Alexis Biochemicals</a>	Reagents for molecular- and cell-biology research	Lausanne, Switzerland	<a href="http://www.alexis-corp.com">www.alexis-corp.com</a>
<a href="#">Attagene</a>	Transcription-factor profiling system; software	Research Triangle Park, North Carolina	<a href="http://www.attagene.com">www.attagene.com</a>
<a href="#">BD Biosciences</a>	Research reagents, bioimaging systems, instrumentations	San Diego, California	<a href="http://www.bd.com">www.bd.com</a>
<a href="#">Biomol</a>	Services for chemical synthesis, cell culture and antibody production	Hamburg, Germany	<a href="http://www.biomol.de">www.biomol.de</a>
<a href="#">Bio-Rad</a>	Products, instruments and software for life-sciences research	Hercules, California	<a href="http://www.bio-rad.com">www.bio-rad.com</a> ●
<a href="#">BMG LABTECH</a>	Microplate and array readers and handling systems	Offenburg, Germany	<a href="http://www.bmglabtech.com">www.bmglabtech.com</a>
<a href="#">Brinkmann Instruments</a>	Laboratory instruments; consumables	Westbury, New York	<a href="http://www.brinkmann.com">www.brinkmann.com</a>
<a href="#">Cambrex</a>	Products for molecular- and cell-biology research	East Rutherford, New Jersey	<a href="http://www.cambrex.com">www.cambrex.com</a>
<a href="#">Cole-Parmer</a>	Instruments and reagents	Vernon Hills, Illinois	<a href="http://www.coleparmer.com">www.coleparmer.com</a>
<a href="#">EMD</a>	Calbiochem, Novabiochem and Novagen product lines	San Diego, California	<a href="http://www.emdbiosciences.com">www.emdbiosciences.com</a>
<a href="#">Enzo Life Sciences</a>	Consumables and assays for molecular biology, gene expression and genomic analysis	New York, New York	<a href="http://www.enzo.com">www.enzo.com</a>
<a href="#">Geneservice</a>	Genomic and proteomic resources; contract services for DNA sequencing, microarray analysis and SNP genotyping	Cambridge, UK	<a href="http://www.geneservice.co.uk">www.geneservice.co.uk</a>
<a href="#">Harvard Apparatus</a>	Instruments and equipment for electrophysiology and cell biology	Holliston, Massachusetts	<a href="http://www.harvardapparatus.com">www.harvardapparatus.com</a>
<a href="#">Horiba Jobin Yvon</a>	Spectroscopy systems and accessories including Raman, atomic emission, and UV spectroscopy	Edison, New Jersey	<a href="http://www.jobinyvon.com">www.jobinyvon.com</a>
<a href="#">INTEGRA Biosciences</a>	Equipment for sterilization, liquid handling, cell culture and sample storage	Baar, Switzerland	<a href="http://www.integra-biosciences.com">www.integra-biosciences.com</a>
<a href="#">Irvine Scientific</a>	Defined media for cell-culture applications; custom media services	Santa Ana, California	<a href="http://www.irvinesci.com">www.irvinesci.com</a>
<a href="#">Lonza</a>	Molecular biology reagents and systems; advanced chemical synthesis	Basel, Switzerland	<a href="http://www.lonza.com">www.lonza.com</a> ●
<a href="#">Merck</a>	Chemicals, kits and reagents for chemistry, molecular and cell biology-related research	Darmstadt, Germany	<a href="http://www.merck.de">www.merck.de</a>
<a href="#">Molecular Devices</a>	Liquid-handling and microplate processing equipment; imaging instruments	Sunnyvale, California	<a href="http://www.moleculardevices.com">www.moleculardevices.com</a> ●
<a href="#">MP Biomedicals</a>	Reagents and chemicals for research	Aurora, Ohio	<a href="http://www.mpbio.com">www.mpbio.com</a>
<a href="#">New England Biolabs</a>	Molecular-biology-related reagents, kits and enzymes	Ipswich, Massachusetts	<a href="http://www.neb.com">www.neb.com</a>
<a href="#">Nalge Nunc International</a>	Labware	Rochester, New York	<a href="http://www.nalgenunc.com">www.nalgenunc.com</a>
<a href="#">PerkinElmer</a>	Instruments, reagents and kits for life sciences	Waltham, Massachusetts	<a href="http://las.perkinelmer.com">las.perkinelmer.com</a>
<a href="#">Promega</a>	Chemicals for mass spectrometry analysis	Madison, Wisconsin	<a href="http://www.promega.com">www.promega.com</a> ●
<a href="#">PREMIER Biosoft International</a>	Software for life-sciences research	Palo Alto, California	<a href="http://www.premierbiosoft.com">www.premierbiosoft.com</a>
<a href="#">Princeton Separations</a>	DNA purification columns and reagents, fluorescent protein labelling kits,	Adelphia, New Jersey	<a href="http://www.prinsep.com">www.prinsep.com</a>
<a href="#">Stratagene</a>	Tools and reagents for molecular biology, genomics and proteomics	La Jolla, California	<a href="http://www.stratagene.com">www.stratagene.com</a>
<a href="#">Takara Bio</a>	Reagents, kits and consumables for molecular biology	Shiga, Japan	<a href="http://www.takara-bio.com">www.takara-bio.com</a>
<a href="#">Tocris Bioscience</a>	Chemicals for life-science research; contract research services	Avonmouth, UK	<a href="http://www.tocris.com">www.tocris.com</a>
<a href="#">Wako Chemicals USA</a>	Speciality chemicals supplier; clinical diagnostic reagents	Richmond, Virginia	<a href="http://www.wakousa.com">www.wakousa.com</a> ●
<a href="#">USB</a>	Chemicals and reagents for molecular biology	Cleveland, Ohio	<a href="http://www.usbweb.com">www.usbweb.com</a>

● see advertisement

# naturejobs

**THE CAREERS  
MAGAZINE FOR  
SCIENTISTS**

**T**he decline in the number of young junior faculty members at US universities could make it difficult to fill leadership positions in the future, according to a study by the American Council on Education. Just 3% of tenure-track faculty members at institutions that run full undergraduate courses are aged 34 or younger, so finding experienced people to be chairs, deans and university presidents will be difficult in the next generation, according to the council's report, *Too many Rungs on the Ladder? Faculty Demographics and the Future Leadership of Higher Education*.

Higher education has been affected by three main issues. First, faculty members are now generally older than they were when mandatory retirement was abolished in 1994, and are reluctant to retire. Second, universities rely more on part-time and non-tenure-track positions. And third, students are completing their doctorates later in life. The trend towards more and longer postdoc courses means that people have less time to gain the experience needed to advance into leadership positions.

The situation has both positive and negative aspects. On one hand, young faculty will have less competition and more positions available when they finally manage to move through the ranks. On the other hand, they might not have the chance to do so until they reach retirement age. So, a professor who is currently 45 years old could well still have the chance to become a dean — or even a university president — but perhaps not until he or she reaches the age of 70.

There are no easy solutions. Universities are unlikely to force ageing faculty off the top rung of the ladder or to create new positions for young professors. However, they may be forced to remove some rungs from the ladder — or else place some closer together, so that today's junior faculty members can progress from chair to dean more quickly. Perhaps today's administration can help mentor tomorrow's university leaders — or, at least, convince them that such positions are worth the wait.

**Paul Smaglik moderates the Naturejobs career forum on Nature Networks.**

## CONTACTS

**Editor:** Gene Russo

**European Head Office, London**  
The Macmillan Building,  
4 Crinan Street, London N1 9XW, UK  
Tel: +44 (0) 20 7843 4961  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**European Sales Manager:**  
Andy Douglas (4975)  
e-mail: [a.douglas@nature.com](mailto:a.douglas@nature.com)

**Natureevents:**  
Ghizlaine Ababou (+44 (0) 20 7014 4015)  
e-mail: [g.ababou@nature.com](mailto:g.ababou@nature.com)

**UK Corporate:**  
Nils Moeller (4953)

**Southwest UK/RoW:**  
Alexander Ranken (4944)

**Northeast UK/Ireland:**  
Matthew Ward (+44 (0) 20 7014 4059)

**France/Switzerland/Belgium:**  
Muriel Lestringuez (4994)

**Scandinavia/Spain/Portugal/Italy:**

Evelina Rubio-Hakansson (4973)

**North Germany/The Netherlands/Eastern**

**Europe:** Reya Silao (4970)

**South Germany/Austria:**

Hildi Rowland (+44 (0) 20 7014 4084)

**Advertising Production Manager:**

Stephen Russell

To send materials use London address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**Naturejobs web development:** Tom Hancock

**Naturejobs online production:** Dennis Chu

**US Head Office, New York**

75 Varick Street, 9th Floor,

New York, NY 10013-1917

Tel: +1 800 989 7718

Fax: +1 800 989 7103

e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

**US Sales Manager:** Peter Bless

**India**

Vikas Chawla (+91 1242881057)

e-mail: [v.chawla@nature.com](mailto:v.chawla@nature.com)

**Japan Head Office, Tokyo**

Chiyoda Building, 2-37 Ichigayatamachi,

Shinjuku-ku, Tokyo 162-0843

Tel: +81 3 3267 8751

Fax: +81 3 3267 8746

**Asia-Pacific Sales Manager:**

Ayako Watanabe (+81 3 3267 8765)

e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)

**Business Development Manager, Greater**

**China/Singapore:**

Gloria To (+852 2811 7191)

e-mail: [g.to@natureasia.com](mailto:g.to@natureasia.com)



# MOVERS

**Stephen Brandt, Director, Oregon Sea Grant, Corvallis, Oregon**



**1997-2008** Director, NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan  
**1994-97** Director, Great Lakes Center for Environmental Research and Education and Professor of Biology, State University of New York College at Buffalo, New York

As the new director of Oregon Sea Grant, Stephen Brandt has eagerly accepted a daunting task: helping the US Pacific coastal regions address fisheries declines and prepare for climate change. It's his latest interaction with Sea Grant, the coastal science programme of the US National Oceanic and Atmospheric Organization (NOAA), which has been a staple of his career since he was conducting graduate research.

California Sea Grant director Russ Moll says that Brandt's background will boost ecosystem-based management efforts. "Stephen is one of those rare folks with the skills to look at the big picture in oceans — which we need as we struggle with ecosystem-wide concerns such as ocean acidification," says Moll.

Brandt started his science career with a mathematics degree at the University of Wisconsin in Madison. But the outdoorsman decided to get a second degree in zoology, and spent weekends conducting field work on Wisconsin's freshwater lakes. That led to a graduate project applying sonar to study fish dynamics, then a PhD using underwater acoustics to see how temperature affects habitat preference in Great Lakes fish. But instead of accepting a tenure-track position there, he joined the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Marine Laboratories in Australia. "I sought adventure when, at 28, I took the Australia position — and I got a full-blooded marine experience," he says.

At the CSIRO, he refined acoustic approaches to investigate how Australia's vast, warm eddies might serve as nursery grounds for fish in the open sea. After four years, Brandt returned to the United States to study the Great Lakes' evolving salmon fishery with Sea Grant's programme at the State University of New York in Syracuse. Later, he studied the largest US estuary at the University of Maryland's Chesapeake Biological Laboratory.

When Sea Grant's Great Lakes Center for Environmental Research and Education was created in 1994, Brandt jumped at the chance to direct it. Four years later, he was overseeing the NOAA's Great Lakes Environmental Research Laboratory in Ann Arbor, Michigan, where he created a single 'science' branch to strengthen the cross-disciplinary work that bolsters their now-leading role in ecosystem forecasting. Moll says that Brandt's past success with region-wide projects will help the west coast to tackle the effects of climate change, including organism range shifts and increased storminess.

**Virginia Gewin**

## NETWORKS & SUPPORT

### Masters of professional science

Some people have expressed concern about the legitimacy, usefulness and costs of professional science master's (PSM) degrees, a relatively new US entity that could promise additional career avenues for fledgling scientists (see *Nature* **454**, 547; 2008). As advocates and purveyors of the PSM, we would like to address those concerns and endorse it as an option for scientists seeking management and science training in just a few years.

First, it is not intended for students considering a PhD, although a few PSM graduates do continue on to the PhD after becoming excited by the research to which they have been exposed. Rather, it is designed for students and science professionals who want to work in non-academic sectors, in interdisciplinary fields and in emerging areas. Science professionals looking to gain a competitive edge, re-enter the workforce or refine their skills may also find it worthwhile.

Second, the PSM is a relatively new degree; there are approximately 2,100 graduates nationwide. Not enough data exist yet to declare it a clear success. But the data we have are promising. A recent survey showed that in two years alone, the number of programmes grew by at least 20% (2006-08), and enrolment increased

54% (2004-06). Placement data are also encouraging: almost 70% of 2006 graduates who were not already working full-time found employment in business, government and non-profit sectors, either before they graduated or immediately after.

Cost is an issue. In contrast to PhDs, the master's degree is usually funded by the student. However, the highly competitive salaries PSM graduates can expect make it a worthwhile investment. A recent report from the National Research Council (NRC) notes a strong and growing current demand for master's-level science professionals and healthy growth in the salaries of master's degree-holders in science and engineering — salaries that have grown faster during the past ten years than those of PhD holders. The NRC advocated financial aid for PSM students.

There is good evidence to date that the PSM is a worthwhile investment, and that it benefits the institution and the employer. We are confident that forthcoming data will support the PSM even more strongly.

**Eleanor Babco is co-project director, Professional Master's Initiatives. Carol Lynch is senior scholar in residence and director of Professional Master's Programs at the Council for Graduate Schools.**

#### POSTDOC JOURNAL

### The coming challenge

In 2009, I will start yet another project: a baby. It is a terrifying prospect. But many seem surprised at my financial anxiety, given the Singapore government's policies aimed at boosting a birthrate in decline. Incentives range from financial bonuses to the creation of a fund to encourage family-friendly work practices. Some even say that childbirth is a woman's 'national service' or duty.

The tax breaks and additional days of childcare leave provide welcome relief, but as a researcher, many family-friendly options are not feasible, such as extended maternity leave or working part-time or from home. Given the high expectations of employers, a career break might mean career suicide. We are evaluated according to productivity, which is inevitably affected by parenthood. One non-scientist relative of mine was told that her maternity leave cost her a promotion. The competitive environment may be exacerbated by single people and childless couples who are upset by policies they perceive as discriminatory.

A newspaper article here recently profiled two successful female senior researchers, citing them as role models. One is single, the other divorced. Being successful may come at the cost of one's marriage. As I prepare to start a family, I must re-evaluate my priorities. I will soon discover for myself how Singapore's biomedical research community defines 'work-life balance'.

**Amanda Goh is a postdoctoral fellow in cell biology under the Agency of Science, Technology and Research in Singapore.**

naturejobs

More than  
just jobs!

Movers

Careers

Focuses

Regions

Podcasts

Prospects

Spotlights

Career fairs

Job channels

Announcements

Special reports

Visit

naturejobs.com

nature publishing group National University of Ireland, Galway  
*Ollscoil na hÉireann, Gaillimh*

## ESTABLISHED PROFESSOR OF PLANT SCIENCE

The National University of Ireland Galway is a student-centred, research intensive University of international standing, with a staff committed to excellence. The University is committed to delivering a stimulating learning environment and providing leadership in key scientific, economic, social, and cultural areas.

To further enhance the University's mission, applications are now invited for the post of Established Professor of Plant Science within the School of Natural Sciences at NUI Galway. The holder of the Chair will provide academic leadership to the Discipline of Botany/Plant Science within the School.

The Discipline of Botany/Plant Science is a key component of a newly established School, within the College of Science, that also includes Biochemistry, Earth & Ocean Science, Microbiology and Zoology. The various Disciplines within the School have a long-established international reputation for both teaching and research, led by more than 50 permanent members of academic staff. The School offers four-year BSc (Honours) undenominated degree programmes in each of its five cognate disciplines, as well as in several interdisciplinary, denominated areas. Its members are also active in taught and research Graduate programmes that lead to MSc and PhD Degrees, as well as Higher Diplomas.

All current academic members of staff in the Discipline of Botany/Plant Science are actively engaged in scholarship and research. Within the School, research in biological and environmental sciences span all levels of organisation, from molecular and cell biology, through research at the organismal level, to the study of populations, communities and ecosystems. These activities are aligned with the University's strategic, cross-cutting research areas in biomedical research, energy, environment and the marine. Members of the School and Discipline have strong interactions with the University's Research Institutes, including the Environmental Change Institute, the National Centre for Biomedical Engineering Science and the Martin Ryan Marine Science Institute and also with other Schools across the University. Further details can be obtained in the supporting documentation and web sites listed therein.

Applicants for the Chair should have relevant expertise in Molecular/Cell Science or in the area of the Environment (Terrestrial, Freshwater or Marine). It will also be an advantage if a candidate's interests are a good fit to the strategic goals of the School and University and further strengthen interactions with some of the University's cognate Research Institutes.

*The successful candidate shall possess:*

- The capacity to provide dynamic academic leadership in the development of the discipline and generally in the promotion of teaching and research in the plant sciences
- An internationally-recognised track record for high-quality research in a relevant area as evidenced by an exceptional publication record in top-ranking, peer-reviewed international journals, a proven ability to develop innovative research together with demonstrated success in competing for funding at national and international levels
- A strong commitment to excellence in teaching and learning at all levels within the University as evidenced by significant experience of course delivery and design
- The capacity to represent effectively the Discipline and the School, inside and outside the University
- The ability to potentially act as Head of School and, in particular, demonstrated excellent interpersonal, communication and other relevant skills appropriate to that role
- A willingness to participate in the overall life of the University.

**Additional information** is available at: <http://www.nuigalway.ie/botany>

**For informal discussion contact:** Dr. Gerry Morgan, Dean of the College of Science, 091-493615 or email [dean.science@nuigalway.ie](mailto:dean.science@nuigalway.ie)

W171017R

[www.nuigalway.ie](http://www.nuigalway.ie)



science foundation ireland  
fondúireacht eolaíochta éireann

## Science Foundation Ireland

**Ireland**  
funds great research...  
...maybe it's your turn!

### Science Foundation Ireland, (SFI)

the national foundation for excellence in scientific research is investing in academic researchers and research teams who are most likely to generate new knowledge, leading edge technologies, and competitive enterprises.

SFI has a flexible grants and awards portfolio and several times a year issues calls for proposals from scientists and engineers. SFI's award programmes include:

#### Principal Investigator Programme

for outstanding researchers, normally ranging between €50,000 - €1 million per year and may be up to five years in duration.

#### Research Professor Recruitment Awards

for outstanding researchers, with particularly distinguished international reputations, awards normally ranging up to €500,000 per annum for up to two years.

#### E.T.S. Walton Visitor Awards

supporting leading international scientists who visit Ireland to undertake research for up to one year, normally ranging up to €200,000. **Call opens 6 October 2008**

#### President of Ireland Young Researcher Awards (PIYRA)

attracting to Ireland and supporting Irish researchers within five years of completing their PhD, normally up to €1 million over five years.

### Career Opportunities at SFI

SFI's culture is innovative, highly motivated, and dynamic. We employ people from science, business, technology, and academic backgrounds who can help strengthen an outstanding Irish research environment.

For details on current employment opportunities at SFI please visit [www.sfi.ie](http://www.sfi.ie) or contact [hr@sfi.ie](mailto:hr@sfi.ie)

### Current employment opportunities on SFI funded research projects

#### University College Dublin

Further Details on UCD Posts:

[www.ucd.ie/jobopportunities](http://www.ucd.ie/jobopportunities)

**UCD School of Electrical, Electronic and Mechanical Engineering**  
**UCD Engineering and Materials Science Centre**

Charles Parsons Award Researcher - Fixed Term 6 Year Post

**UCD School of Agriculture, Food Science & Veterinary Medicine**  
**Reproductive Biology Research Cluster (RBRC)**

Postdoctoral Fellow – Fixed Term Post

**UCD School of Biomolecular and Biomedical Science**

Postdoctoral Fellow - Cell and Molecular Biology - Fixed Term Post

Postdoctoral Fellow - Molecular Biology & Bioinformatics - Fixed Term Post

Further Details on UCD Studentships:

[www.ucd.ie/graduatestudies/opportunities](http://www.ucd.ie/graduatestudies/opportunities)

**UCD School of Medicine and Medical Sciences**  
**Conway Institute of Biomolecular and Biomedical Research**

Phd Studentships to Study the Contribution of the IL-13 Receptor Complex to the Development of Pulmonary Fibrosis

**UCD School of Agriculture, Food Science & Veterinary Medicine**  
**Reproductive Biology Research Cluster (RBRC)**

Phd Projects available include the Regulation of MHC-1 Expression during Preimplantation Embryo Development and Molecular Analysis of Mammalian Oocyte Maturation

Science Foundation Ireland

[www.sfi.ie](http://www.sfi.ie)



**Tyndall National Institute**Further Details: [www.tyndall.ie/careers](http://www.tyndall.ie/careers)

Phd position - Theory of Next Generation Epitaxial Quantum Dot Materials and Devices

Phd position - Design of Molecular and Nanoscale Electronic Devices

Phd position - In-situ Spectroscopic Reaction Monitoring

Phd position - Design and Synthesis of Novel Precursors

Phd position - Kinetic Monte Carlo Simulations

Phd position - Electrical Characterisation of Dielectric/Gate Structures

Phd position - Inelastic Electron Tunneling Spectroscopy

Phd position - First Principles Modelling

Phd position - Nano-casting Routes

Phd position - Interconnect Materials

Phd position - Advanced Thin Film Growth by CVD and ALD

Phd position - Development of Large-Scale Colloidal Crystallisation Methods for the Production of Photonic Crystals

Post-doc position - Experimental Physics of Photonic Materials and Devices

Post-doc position - Theory of Extreme Semiconductor Alloys

Phd position - Avalanche Photodetectors

Phd position - Miniaturised Systems for the Built Environment

Phd position - Develop energy harvesting mechanisms for the next generation, wearable wireless sensor network systems

Phd position - Investigate &amp; Develop Photonics Systems Research

Phd position - Long Wavelength, Site Controlled Pyramidal Quantum Wires

Phd position - Biology of novel vascular progenitor cells and the implications for this field by developing platform diagnostic devices for cardiovascular disease

Phd position - Solid State Electronic Structure Theory

Phd position - Development of Electrical Contacting Strategies for Template-Based Nanostructures and Nanostructured Arrays

Phd position - Synthesis and Characterisation of Optically Active Nanostructured Oxides

Phd position - Harmonic Mode Locking in MultiWavelength Fabry-Perot Lasers

Phd position - Nonlinear Optics of Refractive Index Profiles with Tailored Group Velocity

Post-doc position - Optical Signal Processing for High Capacity Networks

Post-doc position - All-Optical Switching Techniques for Optical Networks

Post-doc position - Wearable Wireless Sensor Network Systems for Health

Post-doc position - Energy Aware Hardware for Wireless Sensor Network Systems

Post-doc position - Deployment Engineer for Wireless Systems

Post-doc position - Advanced Semiconductor Laser, Simulation, Design, Characterisation and Test

Post-doc position - Packaging Engineer

Post-doc position - Electron Microscopy Technician

Post-doc position - Generation and Characterisation of Ferroelectric and High-K Metal Oxide Dielectric Nanostructures

Senior Post-doc position - Photonic Communication System Research

Post-doc position - Energy Harvesting for Wireless Sensors in the Built Environment

**National University of Ireland, Maynooth**Further Details: [www.cs.nuim.ie/~tnaughton/](http://www.cs.nuim.ie/~tnaughton/)

Post-doc position (12 months) Department of Computer Science - Image processing and analysis of digital holograms of real-world three-dimensional objects

**Trinity College Dublin**Further Details: [www.tcd.ie/vacancies](http://www.tcd.ie/vacancies)**School of Genetics and Microbiology  
Smurfit Institute of Genetics**

2 Postdoctoral Fellowships in Bioinformatics and Genome Evolution

Postdoctoral Fellowship

PhD Studentship

**School of Medicine  
Department of Clinical Medicine**

Phd Studentship (3-year contract)

**School of Engineering  
Trinity Centre for Bioengineering**

Phd Studentship in Computation Studies in Mechanobiology

**CRANN (Centre for Research on  
Adaptive  
Nanostructures and  
Nanodevices)**

Postdoctoral Researcher (1-year contract)

**National University of Ireland,  
Galway**

Further Details:

[www.nuigalway.ie/vacancies](http://www.nuigalway.ie/vacancies)

Postdoctoral Researcher in Nuclear Organisation and Gene Expression

**Dublin City University**

Further Details:

[www.dcu.ie/vacancies/current.shtml](http://www.dcu.ie/vacancies/current.shtml)

Postdoctoral Research in Optical Communications and High Speed Opto-Electronics - 3 year contract RINCE

**Biomedical Diagnostics Institute  
(BDI)**

Further Details:

[www.bdi.ie/about\\_bdi/careers.html](http://www.bdi.ie/about_bdi/careers.html)

Postdoctoral Researcher - Microfluidics &amp; Polymer Microfabrication - 12 month contract

Postdoctoral Researcher - Surface Science of Bioassay Devices - 12 month contract

Research Assistant (position based at Royal College of Surgeons in Ireland) - Immunoassay/Protein Biochemist - 12 month contract

**Science Foundation Ireland**Wilton Park House, Wilton Place, Dublin 2, Ireland. tel +353 1 607 3200 fax +353 1 607 3201 email [info@sfi.ie](mailto:info@sfi.ie)Apply for an SFI award or learn more about our programmes at [www.sfi.ie](http://www.sfi.ie)

## Professorships and Lectureships in Science, Engineering and Computing

Dublin City University ([www.dcu.ie](http://www.dcu.ie)) is Ireland's youngest university and our rapid growth and leadership has helped to transform many of the traditional practices and assumptions of higher education. Through our current strategic plan we aim to distinguish ourselves through our engagement with industry and business in building internationally competitive research teams, which will succeed in converting the results of high quality basic research into significant social and economic impacts.

DCU wishes to invite **Expressions of Interest** from high quality candidates at both **Lecturer and Professor** levels, in context of the preparation of the DCU submission to the Science Foundation Ireland 'Stokes Professorship and Lecturer Programme' 2009 ([www.sfi.ie](http://www.sfi.ie)), with a specialty in one or more of the following fields:

### Biological/Health Sciences

- Biomedical Diagnostics
- Drug Discovery and Development
- Cancer Research
- Therapeutics

### Engineering

- Image processing & Analysis Network Innovations
- High Speed Devices and Systems and
- Content Analysis in Context
- Innovative Manufacturing
- Biomedical Engineering Sustainable Technology

### Computing

- Software Engineering, Language Technologies, Scientific
- Computing Information Security

### Physical/Chemical Sciences

- Photonic Sciences
- Plasma Science and Technology
- Environmental Monitoring

Further details on the application process are available at: [www.dcu.ie/vacancies/current.shtml](http://www.dcu.ie/vacancies/current.shtml)

**Expressions of Interest should be emailed to: [hr.applications@dcu.ie](mailto:hr.applications@dcu.ie) by 5pm on Friday October 24th.**

W171220R

**Is your work-life  
out of balance?**



**Keep up-to-date with all  
the latest issues affecting  
scientists and their  
working environment on  
[naturejobs.com](http://naturejobs.com)**

nature publishing group **npg**

# Mars is the wrong colour

No longer seeing red.

**Ian Randal Strock**

There's a difference between a conspiracy and what Margaret Mead called "a small group of thoughtful, committed people". The former will always fail, because someone will give it away. The latter, as Mead said, "can change the world. Indeed, it is the only thing that ever has."

Many ventures fail not for a bad idea or poor planning, but because the wrong group of people is involved. I'd learned that time and again in my organizing efforts: fan groups, idealistic businesses and even social organizations that sputtered along but eventually collapsed because the people involved didn't mesh well, or didn't have the right skill sets.

This time, I vowed, it would be different. This time, my goal dwarfed all the others, and required far more commitment and cohesion among the people I'd be gathering to carry out the plan. So I went slowly. I didn't start with public pronouncements or marketing campaigns or mass appeals to everyone I knew.

Instead, I found people of like mind through quiet, one-on-one conversations. I found my potential collaborators at scientific conferences, through friends and business acquaintances and, in one instance, a call out of the blue to an author who wrote something in a story that resonated with me.

It was a large group for a conspiracy, but a small number of people to carry out such a grandiose plan. And this one time, finally, I knew before I asked that they would all say yes. I convinced my double handful of potential co-conspirators to get together at a quiet resort during the off-season. They all said yes to the first date I proposed.

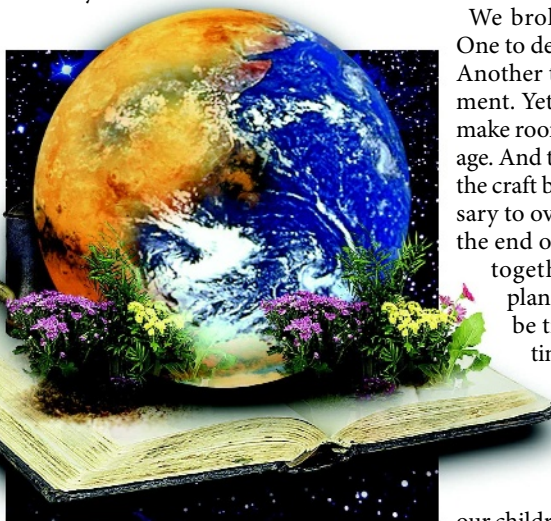
After the introductions were out of the way, I think they all realized I had something larger in mind than a social gathering. I stood up, and everyone else fell silent, looking at me with anticipation.

"Mars is the wrong colour," I said. "After so many missions looking for life, I think we can be fairly certain that there is none on Mars. There may once have been, but there's nothing now, and the environment won't permit anything we'd recognize as life to grow."

There were nods of agreement around the room, and an encouraging, "Yes, go on."

"We can keep going as we have been, sending probes to Mars every two years to investigate smaller and smaller possibilities, or we can gather a small group of people," I looked around the room with purpose, and knew they were the right crowd, because every one of them met my eyes and smiled, "and admit there's no life, but that it's time to start seeding it."

"NASA is launching the Firebird in a year. It will be a successor to the Phoenix, with a soil laboratory, a digging arm and mobility."



It was definitely the right crowd. I didn't have to finish the concept before someone else said: "The Flora experiment. We'll need to co-opt it, make it smaller and lighter than the programme integrators expect."

"Then we can take the extra space and put what in it?"

"Extremophiles. Some sort of microbes that will be happy in the current Martian environment, and which will excrete —"

"The oxygen and ozone we need to transform the environment."

"Someone will notice."

"By then, it'll be too late to stop the process. If word doesn't get out, the rest of the world may assume it's a natural process."

"Or a miracle."

"But we'll know better."

"We won't have a living Mars to visit, nor will our grandchildren."

"No, but *their* grandchildren will. Governments move too slowly, and don't think far enough ahead."

"Better a second habitable planet in a bunch of generations than none at all."

"If we're caught..."

"What? No country I know has a law against terraforming Mars."

"But they'll come up with something. Theft of government services? Deceit? Lying to a federal agency?"

"Then we'll pay a fine or go to jail."

"But Mars will live."

"What do we call the project?"

"If there's a name, there will be something to give us away. Something to let slip."

"Don't use your government e-mail accounts to talk about this with anyone."

We broke up into working groups. One to decide which microbes to send. Another to build their flight compartment. Yet another to redesign Flora to make room for our unannounced package. And those who would be inspecting the craft before launch, who were necessary to overlook our modifications. By the end of the weekend, we came back together, and we had a workable plan. The scheduling was going to be tight, but that only meant less time for our plot to be discovered and stopped.

We all left that weekend vowing to carry out our plan without telling a soul: not our spouses, not our children, and leave not a word in our wills. What we were doing, we were doing for all humanity.

I organized the group: that was my contribution. I'm not a scientist, not a biologist, I don't work for NASA, and I had nothing more to do with any of the project members. But ego is a funny thing. As the organizer, I feel a certain responsibility to them, to let the world know that there is hope; the hope of another planet for our grandchildren's grandchildren, if only we can keep ourselves alive long enough for Mars to be ready. Firebird will land on Mars very soon; there's nothing anyone can do now to stop the process, short of nuking Mars, but that's not going to happen. So I've written this story.

Keep looking up, and tell your children. There will come a time when Mars is no longer the red planet. Watch for the white clouds, the blue water and the green life. It's coming.

**Ian Randal Strock** (<http://ianrandalstrock.livejournal.com>) is the editor of **SFScope.com**, the online trade journal of the science-fiction fields. He is also the author of *The Presidential Book of Lists* (Random House/Villard, October 2008).

JACEY